Algorithm Design and Analysis for Large-Scale Semidefinite Programming and Nonlinear Programming

A Thesis
Presented to
The Academic Faculty

by

Zhaosong Lu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
August 2005

Algorithm Design and Analysis for Large-Scale Semidefinite Programming and Nonlinear Programming

Approved by:

Dr. Renato D.C. Monteiro, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Arkadi S. Nemirovski, Co-advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Alexander Shapiro
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Earl R. Barnes
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. William L. Green
School of Mathematics
Georgia Institute of Technology

Date Approved: June 21, 2005

This thesis is dedicated to my parents and my wife Yanchao. Throughout the course of my studies, they have given me much love, support and happiness, each in their unique way.

ACKNOWLEDGEMENTS

During my time at Georgia Tech, there have been many people around who have helped make my studies a truly enjoyable part of my life. In the next few paragraphs, I would like to point out a few of these people to which I am especially in debt.

First and foremost, I would like to thank my advisor, Renato Monteiro, who has spent countless hours over the past five years guiding me in my academic pursuits. Without his clear thinking, enthusiasm, understanding, encouragement and faith in me this thesis would never have happened. I also appreciated his treatment of me more as a colleague and friend rather than as just a graduate student.

I would also like to thank Arkadi Nemirovski, my co-advisor, who has spent a lot of time over the past one and a half years guiding me on an interesting research project. His patience, enthusiasm and encouragement made me feel very enjoyable. His deep insight and strong motivation will inspire me through my future career.

Next, I would like to thank Ellis Johnson, who led me into a real-world operations research project. I have learned from him what I could not acquire from classes. His patience and enthusiasm made me have a wonderful time working with him.

I would also like to thank my committee for spending time carefully reading my thesis. Their guidance and advice are invaluable to the completion of this thesis. I would also express my gratitude to Gary Parker, Associate Chair for Graduate Studies, who has been terrific, always answering my many questions and concerns.

Finally, I would like to thank my friends and fellow graduate students who have made life at school and away from school lots of fun. I am especially in debt to my office mate Jerome O'Neal for a lot of exciting collaboration and discussion. Many thanks also to Junxia Chang, Xiaodong Yang, and Yunkai Zhou for their friendships. I hope they each know how much I have valued their friendships over the past few years, and I wish them all the best of luck in the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

SDP	abbreviation for *semidefinite programming* or *semidefinite program*
LP	abbreviation for *linear programming* or *linear program*
NLP	abbreviation for *nonlinear programming* or *nonlinear program*
IPM	abbreviation for *interior-point method*
$\Re$	the set of real numbers
$\Re^p$	the set of real p-length column vectors; the i-th entry of $x \in \Re^p$ is x_i
$\Re^p_+$	the set of entry-wise nonnegative real p-length column vectors
$\Re^p_{++}$	the set of entry-wise positive real p-length column vectors
$\Re^{p\times q}$	the set of real matrices having p rows and q columns; the ij-th entry of $X \in \Re^{p\times q}$ is X_{ij}, the i-th row is $X_{i\cdot}$, and the j-th column is $X_{\cdot j}$
$\mathcal{S}^p$	the set of $p \times p$ real symmetric matrices
$\mathcal{S}^p_+$	the set of $p \times p$ real symmetric positive semidefinite matrices; $X \in \mathcal{S}^n_+$ is equivalently written $X \succeq 0$
$\mathcal{S}^p_{++}$	the set of $p \times p$ real symmetric positive definite matrices; $X \in \mathcal{S}^n_{++}$ is equivalently written $X \succ 0$
$\cdot^T$	for $X \in \Re^{p\times q}$, X^T denotes the matrix transpose of X; also applies to $x \in \Re^p$
$\cdot^*$	for a linear operator Ω, Ω^* denotes its adjoint
$\mathrm{Tr}(\cdot)$	for $X \in \Re^{p\times p}$, $\mathrm{Tr}(X)$ denotes the trace of X, i.e., the sum of the diagonal elements of X
$\cdot \bullet \cdot$	for $A \in \Re^{p\times q}$ and $B \in \Re^{p\times q}$, $A \bullet B$ equals $\mathrm{Tr}(A^T B)$
$\|\cdot\|$	for $x \in \Re^p$, $\|x\|$ denotes the Euclidean norm of x; that is $\|x\| = \sqrt{x^T x}$
$\|\cdot\|_F$	for $X \in \Re^{p\times q}$, $\|X\|_F$ denotes the Frobenius norm of X; that is, $\|X\|_F = \sqrt{X \bullet X}$
$\lvert\cdot\rvert$	for a finite set S, $\lvert S\rvert$ denotes the cardinality of S

$\mathrm{Diag}(\cdot)$	for $x \in \Re^p$, $\mathrm{Diag}(x)$ is the $p \times p$ diagonal matrix having i-th diagonal entry x_i
$\mathrm{diag}(\cdot)$	for $X \in \Re^{p \times p}$, $\mathrm{diag}(X)$ is the p-length column vector having i-th entry X_{ii}; it is the adjoint of $\mathrm{Diag}(\cdot)$
e_i	$e_i \in \Re^p$ has a one in position i and zeros elsewhere
e	$e \in \Re^p$ is the vector of all ones
I	$I \in \Re^{p \times p}$ is the identity matrix
$\cdot^{-1}$	if $x \in \Re^p$ satisfies $x_i \neq 0$ for all i, then $x^{-1} \in \Re^p$ uniquely satisfies $x * x^{-1} = e$; if $X \in \Re^{p \times p}$ is nonsingular, then $X^{-1} \in \Re^{p \times p}$ uniquely satisfies $XX^{-1} = I$

SUMMARY

Semidefinite programming (SDP) is a generalization of linear programming. SDP has numerous applications in various fields, such as statistics, structural design, electrical engineering and combinatorial optimization. Interior-point methods (IPMs) are known as polynomial time methods for solving SDPs, and are favorable for small or medium-sized SDPs. It is well-known that weighted central paths play important role in the design and analysis of IPMs for SDPs. The first topic of this thesis is to study limiting behavior of weighted paths associated with the SDP map $X^{1/2}SX^{1/2}$ and provide applications to error bound analysis and superlinear convergence of a class of primal-dual IPMs.

Although SDPs are polynomially solvable, it is still very challenging to solve large-scale SDPs efficiently. The second topic of this thesis is to provide an approach for solving large-scale well-structured sparse SDPs via a saddle point mirror-prox algorithm with $\mathcal{O}(\epsilon^{-1})$ efficiency by exploiting sparsity structure and reformulating SDPs into smooth convex-concave saddle point problems.

The third topic of this thesis is to develop a long-step primal-dual infeasible path-following algorithm for convex quadratic programming (CQP) whose search directions are computed by means of a preconditioned iterative linear solver. A uniform bound, depending only on the CQP data, on the number of iterations performed by a preconditioned iterative linear solver is established. A polynomial bound on the number of iterations of this method is also obtained.

The last topic of this thesis is to develop an efficient "nearly exact" type of method for solving large-scale "low-rank" trust region subproblems by completely avoiding the computations of Cholesky or partial Cholesky factorizations. We also provide a computational study on this method by applying it to solve some large-scale nonlinear programming problems.

CHAPTER I

INTRODUCTION

1.1 Semidefinite Programming (SDP)

In this section, we introduce and discuss semidefinite programming (SDP) that will be studied in Chapters 2 and 3 of this thesis. In Subsection 1.1.1, we briefly introduce SDP and provide some motivations for two research topics on SDP included in this thesis. We discuss some duality results of SDP and its central path in Subsection 1.1.2.

1.1.1 Preliminary remarks

One form of describing semidefinite programming (SDP) problem is as the problem of minimizing a linear function of a symmetric matrix variable X subject to linear equality constraints on X and the essential constraint that X be positive semidefinite (see (1)). The last constraint is nonlinear and nonsmooth, but convex, so semidefinite programs are convex optimization problems.

Prior to the development of interior-point methods (IPMs) for SDP, there has been a somewhat scattered and slow development but relatively long history of SDP. Very early works in solution stability of linear differential equations and control theory has demonstrated the modeling power of linear matrix inequalities (LMI) (see Boyd et al. [15] for a detailed historic account on this). Early approaches to graph theory involving SDP problems have been proposed in Cullum et al. [25], Donath and Hoffman [30, 31] and Lovász [67]. Also, SDP was studied very early by the nonlinear programming (NLP) community [33, 34, 100, 113, 114, 115]. Despite these early developments, SDP has become a central focus and exciting area in the field of optimization only in the last twelve years or so. A major factor behind this explosive interest in SDP was the development of efficient algorithms for their solution and subsequently the discovery of SDP as a powerful modeling tool. Landmark works by Nesterov and Nemirovski [94, 95, 96] develop a deep and unified

theory of IPMs for solving convex programs based on the notion of self-concordant barrier functions. (See their book [97] for a comprehensive treatment of this subject.) In particular, they and, independently Alizadeh in his breakthrough work [2], showed that primal-only and/or dual-only IPMs for linear programming (LP) can all be extended to SDP. Other primal-only methods have also been proposed by Freund [36] and Jarre [56].

From an algorithmic standpoint, most of the polynomial-time interior-point algorithms for LP have been extended to SDP albeit with much more difficult mathematical analysis (see Todd [122] and Monteiro [79] and references therein). Computationally, IPMs have proven successful in effectively solving small- to medium-sized SDPs. Several applications in convex constrained optimization, control theory and combinatorial optimization, which could not be easily handled before, can now be routinely solved. Many problems can be cast in the form of an SDP. For example, LP, optimizing a convex quadratic form subject to convex quadratic inequality constraints, minimizing the volume of an ellipsoid that covers a given set of points and ellipsoids, maximizing the volume of an ellipsoid that is contained in a polytope, matrix norm minimization, cluster analysis using ellipsoids, and a variety of maximum and minimum eigenvalue problems can be cast as SDPs (see Vandenberghe and Boyd [125]). In addition, SDP has applications in minimum trace factor analysis [11, 33, 34, 113, 127] and optimal experiment design [108] in the area of statistics, and in engineering fields such as structural design [8, 10] and control theory [15]. Another important application of SDP is to the solution of NP-hard combinatorial optimization problems, for which SDPs serve as tractable relaxations (e.g., see Shor [115], Lovász and Schrijver [68], Goemans and Williamson [41], Alizadeh [2], Kamath and Karmarkar [57]) and also as the basis for polynomial-time approximation algorithms with strong performance guarantees (e.g., [41]). In fact, the work [41] was another major factor in the increased interest on SDP, specially by the "approximation algorithms" community. Finally, a long list of SDP applications can be found in the Handbook of Semidefinite Programming [112], the books by Ben-Tal and Nemirovski [9] and Boyd et al. [15], as well as the survey papers by Vandenberghe and Boyd [125], Todd [122], and Monteiro [79].

It is well-known that weighted central paths have played important role in the design

and analysis of high performance IPMs for LPs, monotone linear complementarity problems and SDPs (e.g., see Guler [47], Sturm [121], Kojima et al. [63], Preiß and Stoer [107, 106] and Lu and Monteiro [70, 69]). One topic of this thesis is to study limiting behavior of weighted central paths associated with the SDP map $X^{1/2}SX^{1/2}$ and provide applications to error bound analysis and superlinear convergence of a class of primal-dual IPMs (see Chapter 2).

The SDPs that arise as relaxations of combinatorial optimization problems are typically large-scale, and it has become evident in the past few years that current IPMs, in their most naive form, are largely unsuitable in practice when applied to such SDPs due to an inherent high demand for computer time and storage. At the present level of our knowledge, the only way to process numerically large-scale SDPs seems to use simple first-order methods with computationally cheap iterations. Although all known first-order methods in the large-scale case exhibit slow – sublinear – convergence and thus are unable to produce high-accuracy solutions in realistic time, medium-accuracy solutions are still achievable. Historically, the first SDP algorithm of the latter type was the *spectral bundle* method [54] – a version of the well-known bundle method for nonsmooth convex minimization "tailored" to SDPs. A weak point of the *spectral bundle* method, at least from the theoretical viewpoint, is the convergence rate: the inaccuracy in terms of the objective can decrease with the iteration count t as slowly as $O(t^{-1/2})$ (this is the best possible, in the large scale case, rate of convergence of first-order methods on nonsmooth convex programs). Some other first-order methods based on NLP reformulations were recently proposed in [20, 19]. Theoretical convergence rate results are not established for these methods. Also, as opposed to the primal-dual IPMs, these methods are primal- or dual-only methods and the design of a universal set of termination conditions that works well for all classes of SDP problems is not obvious, and might not be feasible. Recently, novel $O(t^{-1})$-converging first-order algorithms, based on smooth saddle-point reformulation of nonsmooth convex programs were developed [92, 91, 90]. Numerical results presented in these papers demonstrate high computational potential of the proposed methods. However, theoretical and computational advantages exhibited by the $O(t^{-1})$-converging methods as compared to algorithms like spectral bundle

have their price, specifically, the necessity to operate with eigenvalue decompositions of matrices rather than computing a few largest eigenvalues. As a result, the algorithms from [92, 91, 90] as applied to SDP become impractical, when the largest size of diagonal blocks in the matrices exceeds about 1000. Another topic of this thesis is to provide an approach for solving large-scale well-structured sparse SDPs via a saddle point mirror-prox algorithm with $\mathcal{O}(\epsilon^{-1})$ efficiency by exploiting sparsity structure and reformulating them into smooth convex-concave saddle point problems (see Chapter 3).

Before ending this subsection, we introduce some notations that will be used throughout this section. $\mathcal{S}^n$ denotes the space of $n \times n$ symmetric matrices, and $X \succeq 0$ indicates that X is positive semidefinite. We also write $\mathcal{S}^n_+$ for $\{X \in \mathcal{S}^n : X \succeq 0\}$, and $\mathcal{S}^n_{++}$ for its interior, the set of positive definite matrices in $\mathcal{S}^n$. For $X \in \mathcal{S}^n_+$, $X^{1/2}$ denotes its positive semidefinite square root. If $X \in \mathcal{S}^n_{++}$, we write $X^{-1/2}$ for the inverse of $X^{1/2}$, or equivalently the positive semidefinite square root of X^{-1}. We use $A \bullet B$ to denote the inner product $\mathrm{Tr}(A^T B)$ of two $m \times n$ matrices A and B, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. Given a linear operator $\mathcal{F} : E \to F$ between two finite dimensional inner product spaces $(E, \langle \cdot, \cdot \rangle_E)$ and $(F, \langle \cdot, \cdot \rangle_F)$, its *adjoint* is the unique operator $\mathcal{F}^* : F \to E$ satisfying $\langle \mathcal{F}(u), v \rangle_F = \langle u, \mathcal{F}^*(v) \rangle_E$ for all $u \in E$ and $v \in F$. A linear operator $\mathcal{G} : E \to E$ is called *self-adjoint* if $\mathcal{G} = \mathcal{G}^*$. Moreover, $\mathcal{G}$ is said to be *positive semidefinite* (resp. *positive definite*) if $\langle \mathcal{G}(u), u \rangle_E \geq 0$ (resp., $\langle \mathcal{G}(u), u \rangle_E > 0$) for all $0 \neq u \in E$.

1.1.2 The SDP problem, duality and the central path

In this subsection, we introduce the pair of primal-dual SDP problems which will be a subject of our study and discuss some of the duality results that hold for them. We also describe the associated central path which plays an important role in primal-dual interior-point algorithms.

We consider the SDP given in the following standard form:

$$\begin{array}{lll} (P) \qquad \min_X & C \bullet X & \\ & A_i \bullet X = b_i, & i = 1, \ldots, m, \\ & X \succeq 0, & \end{array} \tag{1}$$

where each $A_i \in \mathcal{S}^n$, $b \in \Re^m$, and $C \in \mathcal{S}^n$ are given, and $X \in \mathcal{S}^n$. Throughout this subsection, we assume that the set of matrices $\{A_i\}$ is linearly independent. The dual problem associated with (P) is:

$$\begin{array}{lll} (D) & \max_{y,S} & b^T y \\ & \sum_{i=1}^m y_i A_i \; + \; S & = \; C, \\ & S & \succeq \; 0, \end{array} \tag{2}$$

where $y \in \Re^m$ and $S \in \mathcal{S}^n$. We write $F(P)$ and $F(D)$ for the sets of feasible solutions to (P) and (D) respectively, and correspondingly $F^0(P)$ and $F^0(D)$ for the sets of strictly feasible solutions to (P) and (D) respectively; here "strictly" means that X or S is required to be positive definite. Hence

$$\begin{aligned} F(P) &:= \{X \in \mathcal{S}^n_+ : A_i \bullet X = b_i,\, i = 1, \cdots, m\}, \\ F^0(P) &:= \{X \in F(P) : X \in \mathcal{S}^n_{++}\}, \\ F(D) &:= \{(y, S) \in \Re^m \times \mathcal{S}^n_+ : \sum_{i=1}^m y_i A_i + S = C\}, \\ F^0(D) &:= \{(y, S) \in F(D) : S \in \mathcal{S}^n_{++}\}. \end{aligned}$$

The optimal values for (P) and (D) will be denoted by $\mathrm{val}\,(P)$ and $\mathrm{val}\,(D)$, respectively.

We start by giving the following simple result commonly referred to as the "weak duality lemma".

Proposition 1.1.1 *If X and (y, S) are feasible in (P) and (D) respectively, then*

$$C \bullet X - b^T y = X \bullet S \geq 0.$$

Thus, the quantity $X \bullet S$ is the "excess" of the primal objective function value $C \bullet X$ over the dual value $b^T y$. It is commonly referred to as the *duality gap* at (X, y, S).

Corollary 1.1.2 *Suppose that X and (y, S) are feasible solutions for (P) and (D) respectively, satisfying $X \bullet S = 0$, or equivalently $XS = 0$. Then X and (y, S) are optimal in their respective problems.*

We say that (P) or (D) satisfies strong duality if there exist X and (S, y) satisfying the assumptions of Corollary 1.1.2. Unfortunately, the optimal values of (P) and (D) are not necessarily equal for every SDP problem, and even if they are, strong duality does not necessarily have to hold. The following result gives sufficient conditions for the duality gap between (P) and (D) to be zero and/or for strong duality to hold.

Proposition 1.1.3 *The following statements hold:*

a) If $\text{val}\,(P) > -\infty$ *and* $F^0(P) \neq \emptyset$ *then the set of optimal solutions for* (D) *is nonempty and bounded and* $\text{val}\,(P) = \text{val}\,(D)$*;*

b) If $\text{val}\,(D) < \infty$ *and* $F^0(D) \neq \emptyset$ *then the set of optimal solutions for* (P) *is nonempty and bounded and* $\text{val}\,(P) = \text{val}\,(D)$*;*

c) If $F^0(P) \neq \emptyset$ *and* $F^0(D) \neq \emptyset$ *then the set of optimal solutions of* (P) *and* (D) *are nonempty and bounded and* $\text{val}\,(P) = \text{val}\,(D)$*. (Hence, strong duality is satisfied.)*

A proof of Proposition 1.1.3 can be found for example in Section 4.2 of [97]. By Corollary 1.1.2, it is clear that the conditions below (together with X and S belonging to $\mathcal{S}^n_+$) are sufficient for X and (y, S) to be optimal solutions:

$$\begin{array}{rcl} \sum_{i=1}^m y_i A_i \;+\; S & = & C, \\ A_i \bullet X & = & b_i, \quad \text{for } i = 1, \ldots, m, \\ XS & = & 0. \end{array} \tag{3}$$

To simplify the notation slightly, we introduce the operator $\mathcal{A} : \mathcal{S}^n \to \Re^m$ defined as $(\mathcal{A}X)_i := A_i \bullet X$ for all $X \in \mathcal{S}^n$ and $i = 1, ..., m$. Then the adjoint $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ of this operator is given by $\mathcal{A}^* y = \sum_{i=1}^m y_i A_i$ for every $y \in \Re^m$. Using this notation, we can rewrite the equations above as

$$\begin{array}{rcl} \mathcal{A}^* y \;+\; S & = & C, \\ \mathcal{A}X & = & b, \\ XS & = & 0. \end{array} \tag{4}$$

In both (3) and (4) we could replace the last equation equivalently by $X^{1/2}SX^{1/2} = 0$, or by $S^{1/2}XS^{1/2} = 0$.

The central path is defined as the set of solutions $(X, y, S) = (X(\nu), y(\nu), S(\nu)) \in \mathcal{S}^n_+ \times \Re^m \times \mathcal{S}^n_+$ to

$$\begin{array}{rcl} \mathcal{A}^*y + S & = & C, \\ \mathcal{A}X & = & b, \\ XS & = & \nu I, \end{array} \tag{5}$$

for all $\nu > 0$. Clearly any solution to these equations gives strictly feasible solutions to both (P) and (D), since the last condition implies that X and S are nonsingular, hence positive definite. It turns out that the existence of strictly feasible solutions for both (P) and (D) is sufficient for the existence and uniqueness of solutions to (5) for every positive ν.

The key to the proof of the above result is the analysis of a certain barrier problem associated with (P) whose set of optimality conditions is exactly (5). Consider the following barrier function for the cone of positive semidefinite matrices $\mathcal{S}^n_+$:

$$f(X) := -\ln \mathrm{Det} X. \tag{6}$$

(By convention, we call this a barrier function for $\mathcal{S}^n_+$, even though it is defined only for points in $\mathcal{S}^n_{++}$; it clearly tends to $+\infty$ as X in $\mathcal{S}^n_{++}$ converges to a point on the boundary of $\mathcal{S}^n_+$.) We need to deal with the derivatives of f. The first derivative of f at $X \in \mathcal{S}^n_{++}$ is $f'(X) = -X^{-1}$, in the usual sense that

$$f(X + H) = f(X) + [-X^{-1}] \bullet H + o(\|H\|).$$

For the second derivative, we introduce the useful notation $P \odot Q$ for $n \times n$ matrices P and Q (usually P and Q are symmetric). This is an operator defined from $\mathcal{S}^n$ to $\mathcal{S}^n$ by

$$(P \odot Q)U := \frac{1}{2}(PUQ^T + QUP^T). \tag{7}$$

Then it is not too hard to show that the second derivative of f is $f''(X) = X^{-1} \odot X^{-1}$, in the usual sense that

$$f(X + H) = f(X) + f'(X) \bullet H + \frac{1}{2}[(X^{-1} \odot X^{-1})H] \bullet H + o(\|H\|^2).$$

Note that $f''(X)$ is a self-adjoint and positive definite operator:

$$f''(X)U \bullet V = [X^{-1}UX^{-1}] \bullet V = \mathrm{Tr}(X^{-1}UX^{-1}V) = \mathrm{Tr}(X^{-1}VX^{-1}U) = f''(X)V \bullet U$$

and

$$f''(X)U \bullet U = \mathrm{Tr}(X^{-1}UX^{-1}U) = \mathrm{Tr}((X^{-1/2}UX^{-1/2})^2) = \|X^{-1/2}UX^{-1/2}\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm, the norm induced by the inner product we are using; this quantity is nonnegative, and positive unless $X^{-1/2}UX^{-1/2}$ (and hence U) is zero.

Since f'' is positive definite everywhere, f is strictly convex. Now we consider the primal barrier problem:

$$\begin{array}{lllll} (P_\nu) & \min_X & C \bullet X + \nu f(X) & & \\ & & \mathcal{A}X & = & b, \\ & & X & \in & \mathcal{S}^n_{++}, \end{array}$$

for positive ν. Note that the KKT (or Lagrange) conditions for this problem are necessary and sufficient, since the objective function is convex and the constraints linear (apart from the open set constraint). These optimality conditions can be written as

$$\mathcal{A}X = b, \qquad C - \nu X^{-1} - \mathcal{A}^* y = 0,$$

for $X \in \mathcal{S}^n_{++}$, which can alternatively be expressed as (5) by setting $S := \nu X^{-1} \in \mathcal{S}^n_{++}$.

Let us note that (5) also gives the optimality conditions for the dual barrier problem:

$$\begin{array}{lllllll} (D_\nu) & \max_{y,S} & b^T y & - & \nu f(S) & & \\ & & \mathcal{A}^* y & + & S & = & C, \\ & & & & S & \in & \mathcal{S}^n_{++}. \end{array}$$

The next result shows that the central path is well-defined and is strongly related to the optimal solutions of (P_ν) and (D_ν).

Theorem 1.1.4 *Suppose that both (P) and (D) have strictly feasible solutions. Then, for every $\nu > 0$, there is a unique solution $(X(\nu), y(\nu), S(\nu))$ in $\mathcal{S}^n_{++} \times \Re^m \times \mathcal{S}^n_{++}$ to the central path equations (5). Moreover, for every $\nu > 0$, $X(\nu)$ and $(y(\nu), S(\nu))$ are the unique optimal solutions of the barrier problems (P_ν) and (D_ν), respectively.*

A proof of Theorem 1.1.4 can be found for example in Section 2 of [93]. The above result guarantees the existence and uniqueness of points on the central path, but it does not justify calling it a path. This fact will follow if we show that the equations defining it are differentiable, with a derivative that is square and nonsingular at points on the path. Unfortunately, while the equations of (5) are certainly differentiable, the derivative is not even square since the left-hand side maps $(X, y, S) \in \mathcal{S}^n_{++} \times \Re^m \times \mathcal{S}^n_{++} \subseteq \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ to a point in $\mathcal{S}^n \times \Re^m \times \Re^{n \times n}$; XS is usually not symmetric even if X and S are. We therefore need to change the equations defining the central path. There are many possible approaches, which as we shall see lead to different search directions for our algorithms, but for now we choose a simple one: we replace $XS = \nu I$ by $-\nu X^{-1} + S = 0$. As in our discussion of the barrier function f, the function $X \to -\nu X^{-1}$ is differentiable, with derivative $\nu(X^{-1} \odot X^{-1})$. So the central path is defined by the equations

$$\Phi_P(X, y, S) := \begin{pmatrix} & \mathcal{A}^* y & + & S \\ \mathcal{A}X & & & \\ -\nu X^{-1} & & + & S \end{pmatrix} = \begin{pmatrix} C \\ b \\ 0 \end{pmatrix}, \tag{8}$$

whose derivative is

$$\Phi'_P(X, y, S) := \begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \nu(X^{-1} \odot X^{-1}) & 0 & \mathcal{I} \end{pmatrix}, \tag{9}$$

where $\mathcal{I}$ denotes the identity operator on $\mathcal{S}^n$. We have been rather loose in writing this in matrix form, since the blocks are operators rather than matrices, but the meaning is clear. We want to show that this derivative is nonsingular, and for this it suffices to prove that its null-space is trivial. Since similar equations will occur frequently, let us derive this from a more general result.

Theorem 1.1.5 *Suppose the operators $\mathcal{E}$ and $\mathcal{F}$ map $\mathcal{S}^n$ to itself, and that $\mathcal{E}$ is nonsingular*

and $\mathcal{E}^{-1}\mathcal{F}$ is positive definite. Then the solution to

$$\begin{array}{rcrcl} \mathcal{A}^*\Delta y & + & \Delta S & = & R_d, \\ \mathcal{A}\,\Delta X & & & = & r_p, \\ \mathcal{E}\,\Delta X & + & \mathcal{F}\,\Delta S & = & R_{EF} \end{array} \tag{10}$$

is uniquely given by

$$\begin{array}{rcl} \Delta y & = & (\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*)^{-1}(r_p - \mathcal{A}\mathcal{E}^{-1}(R_{EF} - \mathcal{F}R_d)), \\ \Delta S & = & R_d - \mathcal{A}^*\Delta y, \\ \Delta X & = & \mathcal{E}^{-1}(R_{EF} - \mathcal{F}\Delta S). \end{array} \tag{11}$$

Proof. The formulae for ΔS and ΔX follow directly from the first and third equations. Now substituting for ΔS in the formula for ΔX, and inserting this in the second equation, we obtain after some manipulation

$$(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*)\Delta y = r_p - \mathcal{A}\mathcal{E}^{-1}(R_{EF} - \mathcal{F}R_d).$$

Since $\mathcal{E}^{-1}\mathcal{F}$ is positive definite and the A_i's are linearly independent, the $m\times m$ matrix on the left is positive definite and hence nonsingular. This verifies that Δy is uniquely determined as given, and then so are ΔS and ΔX. Moreover, these values solve the equations. □

In our case, $\mathcal{F}$ is the identity, while $\mathcal{E}$ is $\nu(X^{-1}\odot X^{-1})$ with inverse $\nu^{-1}(X\odot X)$. This is easily seen to be positive definite, in the same way we showed that $f''(X)$ was. Hence the theorem applies, and so the derivative of the function Φ_P is nonsingular on the central path (and throughout $\mathcal{S}^n_{++}\times\Re^m\times\mathcal{S}^n_{++}$); thus the central path is indeed a differentiable path.

By taking the trace of the last equation of (5), we obtain the last part of the following theorem, which summarizes what we have observed.

Theorem 1.1.6 *Assume that both (P) and (D) have strictly feasible solutions. Then the set of solutions to (5) for all positive ν forms a nonempty differentiable path, called the central path. If $(X(\nu), y(\nu), S(\nu))$ solve these equations for a particular positive ν, then $X(\nu)$ is a strictly feasible solution to (P) and $(y(\nu), S(\nu))$ a strictly feasible solution to (D), with duality gap*

$$C\bullet X(\nu) - b^T y(\nu) = X(\nu)\bullet S(\nu) = n\nu. \tag{12}$$

A rigorous proof of Theorem 1.1.6 can be found in Section 2 of [93]. It has been shown that as ν tends to 0, the central path $(X(\nu), y(\nu), S(\nu))$ converges to a specific primal-dual optimal solution of (P) and (D) (see Goldfarb and Scheinberg [42], Kojima et al. [64] and Halicka et al. [52, 51]). Indeed, the primal-dual path-following interior-point methods generate a sequence of iterates closely following the central path and approaching primal-dual optimal set.

1.2 Convex Quadratic Programming (CQP)

In this section, we introduce a pair of primal-dual convex quadratic programming (CQP) problems which will be studied in Chapter 4 of this thesis, and discuss some of the duality results that hold for them. We also describe the associated central path which plays an important role in the design and analysis of primal-dual interior-point algorithms for CQP.

We consider the CQP in standard form:

$$\begin{array}{lrcl} (QP) \qquad \min_x \quad p(x) \equiv & \frac{1}{2}x^T Q x + c^T x & & \\ & Ax & = & b, \\ & x & \geq & 0, \end{array} \tag{13}$$

where the data are $Q \in \mathcal{S}^n_+$, $A \in \Re^{m\times n}$, $b \in \Re^m$, and $c \in \Re^n$, and the decision vector is $x \in \Re^n$.

The dual problem associated with (QP) is:

$$\begin{array}{lrcl} (QD) \qquad \max_{x,y,s} \quad d(x,y,s) \equiv & b^T y - \frac{1}{2}x^T Q x & & \\ & A^T y - Qx + s & = & c, \\ & s & \geq & 0, \end{array} \tag{14}$$

where $y \in \Re^m$ and $s \in \Re^n$. We write $F(QP)$ and $F(QD)$ for the sets of feasible solutions to (QP) and (QD) respectively, and correspondingly $F^0(QP)$ and $F^0(QD)$ for the sets of strictly feasible solutions to (QP) and (QD) respectively; here "strictly" means that x or s is required to have positive components. Hence

$$F(QP) \quad := \quad \{x \in \Re^n_+ : Ax = b\},$$

$$
\begin{aligned}
F^0(QP) &:= \{x \in F(QP) : x \in \Re^n_{++}\}, \\
F(QD) &:= \{(x,y,s) \in \Re^n \times \Re^m \times \Re^n_{++} : A^T y - Qx + s = c\}, \\
F^0(QD) &:= \{(x,y,s) \in F(QD) : s \in \Re^n_{++}\},
\end{aligned}
$$

where $\Re^n_+$ and $\Re^n_{++}$ denote the set of n- vectors having nonnegative, and positive components, respectively. The optimal values for (QP) and (QD) will be denoted by $\mathrm{val}\,(QP)$ and $\mathrm{val}\,(QD)$, respectively.

We start by giving the following simple result commonly referred to as the "weak duality lemma".

Proposition 1.2.1 *If x and (x,y,s) are feasible in (P) and (D) respectively, then*

$$p(x) - d(x,y,s) = x^T s \geq 0.$$

Thus, the quantity $x^T s$ is the "excess" of the primal objective function value $p(x)$ over the dual value $d(x,y,s)$. It is commonly referred to as the *duality gap* at (x,y,s).

Corollary 1.2.2 *Suppose that x and (x,y,s) are feasible solutions for (QP) and (QD) respectively, satisfying $x^T s = 0$. Then x and (x,y,s) are optimal in their respective problems.*

We say that (QP) and (QD) satisfys strong duality if there exist x and (x,y,s) satisfying the assumptions of Corollary 1.2.2. The following proposition shows that strong duality holds for (QP) and (QD).

Proposition 1.2.3 *The following statements hold:*

a) If $\mathrm{val}\,(QP) > -\infty$ and $F(QP) \neq \emptyset$ then the set of optimal solutions for (QP) and (QD) are nonempty and $\mathrm{val}\,(QP) = \mathrm{val}\,(QD)$;

b) If $\mathrm{val}\,(QD) < \infty$ and $F(QD) \neq \emptyset$ then the set of optimal solutions for (QP) and (QD) are nonempty and $\mathrm{val}\,(QP) = \mathrm{val}\,(QD)$;

c) If $F(QP) \neq \emptyset$ and $F(QD) \neq \emptyset$ then the set of optimal solutions of (QP) and (QD) are nonempty and $\mathrm{val}\,(QP) = \mathrm{val}\,(QD)$.

Proof. Under the assumption of statement a), it immediately follows from Proposition 5.2.1 of [13] that the set of optimal solutions for (QD) is nonempty and $\mathrm{val}\,(QP) = \mathrm{val}\,(QD)$. Hence, we obtain that $\mathrm{val}\,(QD) < \infty$ and $F(QD) \neq \emptyset$, which together with Proposition 5.2.1 of [13] implies that the set of optimal solutions for (QP) is nonempty. Therefore, statement a) holds. The proof of statement b) is similar the one of statement a). Statement c) immediately follows from Proposition 1.2.1, and statements a) and b). ∎

By Corollary 1.2.2, it is clear that the conditions below (together with x and s belonging to $\Re^n_+$) are sufficient for x and (x, y, s) to be optimal solutions:

$$\begin{array}{rcrcrcl} A^T y & - & Qx & + & s & = & c, \\ & & Ax & & & = & b, \\ Xs & & & & & = & 0, \end{array} \tag{15}$$

where $X = \mathrm{Diag}(x)$.

The central path is defined as the set of solutions $(x, y, s) = (x(\nu), y(\nu), s(\nu)) \in \Re^n_+ \times \Re^m \times \Re^n_+$ to

$$\begin{array}{rcrcrcl} A^T y & - & Qx & + & s & = & c, \\ & & Ax & & & = & b, \\ Xs & & & & & = & \nu e, \end{array} \tag{16}$$

for all $\nu > 0$, where $e \in \Re^n$ is a vector of all ones. Clearly any solution to these equations gives strictly feasible solutions to both (QP) and (QD), since the last condition implies that x and s are strictly positive. It turns out that the existence of strictly feasible solutions for both (QP) and (QD) is sufficient for the existence and uniqueness of solutions to (16) for every positive ν. The proof of the above result is the analysis of a certain barrier problem associated with (QP) whose set of optimality conditions is exactly (16). Consider the following barrier function for the cone of nonnegative orthant $\Re^n_+$:

$$g(x) := -\sum_{i=1}^{n} \ln x_i. \tag{17}$$

(By convention, we call this a barrier function for $\Re^n_+$, even though it is defined only for points in $\Re^n_{++}$; it clearly tends to $+\infty$ as x in $\Re^n_{++}$ converges to a point on the boundary of $\Re^n_+$.)

Note that g is strictly convex. Now we consider the primal barrier problem:

$$\begin{array}{lrcl} (QP_\nu) & \min_x \; p(x) & + & \nu g(x) \\ & Ax & = & b, \\ & x & > & 0, \end{array}$$

for positive ν. Note that the KKT (or Lagrange) conditions for this problem are necessary and sufficient, since the objective function is convex and the constraints linear. These optimality conditions can be written as

$$Ax = b, \qquad Qx + c - \nu X^{-1}e - A^T y = 0,$$

for $x > 0$, which can alternatively be expressed as (16) by setting $s := \nu X^{-1}e \in \Re^n_{++}$.

Let us note that (16) also gives the optimality conditions for the dual barrier problem:

$$\begin{array}{lrcccccl} (QD_\nu) & \max_{x,y,s} \; d(x,y,s) & - & \nu g(s) & & & & \\ & A^T y & - & Qx & + & s & = & c, \\ & & & & & s & > & 0. \end{array}$$

The next result shows that the central path is well-defined and is strongly related to the optimal solutions of (QP_ν) and (QD_ν).

Theorem 1.2.4 *Suppose that both (QP) and (QD) have strictly feasible solutions. Then, for every $\nu > 0$, there is a unique solution $(x(\nu), y(\nu), s(\nu))$ in $\Re^n_{++} \times \Re^m \times \Re^n_{++}$ to the central path equations (16). Moreover, for every $\nu > 0$, $x(\nu)$ and $(x(\nu), y(\nu), s(\nu))$ are the unique optimal solutions of the barrier problems (QP_ν) and (QD_ν), respectively.*

The proof of Theorem 1.2.4 is similar to the one given in Section 2.3 of [55]. The duality gap of the point $(x(\nu), y(\nu), s(\nu))$ satisfies

$$x(\nu)^T s(\nu) = n\nu,$$

according to (16). It is well-known that $(x(\nu), y(\nu), s(\nu))$ is continuously differentiable for $\nu > 0$. Hence, if $\nu \downarrow 0$ then $x(\nu)$ and $(x(\nu), y(\nu), s(\nu))$ will converge to optimal primal and dual solutions, respectively. In addition, the following property holds for CQP along the central path.

Proposition 1.2.5 *The objective function $p(x(\nu))$ of the primal problem (QP) is monotonically decreasing and the objective function $d(x(\nu), y(\nu), s(\nu))$ of the dual problem (QD) is monotonically increasing if ν decreases.*

A proof of Proposition 1.2.5 can be found in Section 2.3 of [55]. Indeed, the primal-dual path-following interior-point methods generate a sequence of points closely following central path and approaching the primal-dual optimal set of CQP. In Chapter 4, we will develop a long-step primal-dual infeasible path-following algorithm for CQP whose search directions are computed by means of a preconditioned iterative linear solver.

1.3 Trust Region (TR) Methods

Trust region (TR) algorithms are classical methods for solving both convex and nonconvex nonlinear optimization problems. They are known to possess strong convergence properties (see Fletcher [32]). In this section, we briefly introduce trust region methods and provide some motivations for the research done in Chapter 5.

We consider an unconstrained optimization problem

$$\min_x f(x),$$

where $f : \Re^n \to \Re$ is a twice continuously differentiable function and $x \in \Re^n$ is the decision variable. At the current iterate x_k, we assume that a *model function* q_k is constructed whose behavior near x_k is similar to that of the actual objective function f. Usually, a model function q_k has the form:

$$q_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p, \tag{18}$$

where $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$, and $H_k \in \Re^{n \times n}$ is some symmetric matrix. Using the mean value theorem, we have

$$f(x_k + p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p,$$

for some $t \in (0, 1)$. This together with (18) implies that the difference between $q_k(p)$ and $f(x_k + p)$ is $O(\|p\|^2)$, and so the approximate error is small when p is small. To obtain the

next step x_{k+1}, we seek the solution p_k to the TR subproblem

$$\min \{ q_k(p) : \|p\|_{M_k} \leq \Delta_k \} \tag{19}$$

where $\Delta_k > 0$ is a TR radius, M_k is a symmetric positive definite matrix, and the norm $\|\cdot\|_{M_k}$ is defined as

$$\|p\|_{M_k} = \sqrt{p^T M_k p}, \quad \forall p \in \Re^n.$$

Given p_k, we compute the ratio

$$\rho_k = \frac{f_k - f(x_k + p_k)}{q_k(0) - q_k(p_k)},$$

where the numerator is usually referred to as the *actual reduction*, and the denominator is the *predicted reduction*. Based on ρ_k, we use the following strategy to update iterate and trust-region radius: if ρ_k is close to one, we set $x_{k+1} := x_k + p_k$ and increase the TR radius for next TR subproblem (since there is a good agreement between the model q_k and the function f over this step); if ρ_k is positive but not close to zero or one, we set $x_{k+1} := x_k + p_k$ and do not alter the TR radius; if ρ_k is close to zero or negative, we set $x_{k+1} := x_k$, and shrink the TR radius by a certain factor.

For the sake of global convergence of a TR method, an "approximate" solution to the TR subproblem (19) is only required, which achieves at least as much reduction in the model q as the reduction achieved by the so called "Cauchy point" (see for example Moré [88] and Chapter 4 of Nocedal and Wright [98]). There are at least three well-known methods available in the literature for finding an "approximate" solution of the TR subproblem (19) (e.g., *dogleg* method [105], *two-dimensional subspace minimization* method [39], and Steihaug method [118]). Each of them has some specific requirement on the matrix H_k (see Chapter 5 of this thesis). Besides those "approximate" methods, there is a method due to Moré and Sorensen [89], which finds an approximate solution of the TR subproblem (19) in a stronger sense (see (224)). We will refer to this method as a nearly exact (NE) method. Since the NE method of [89] requires repeated computations of Cholesky factorizations of diagonal displacements of H_k, it is suitable only for small- to medium-sized problems.

One topic of this thesis is to develop a method for computing NE solutions of a "low-rank trust region" (LRTR) subproblem whose H_k and M_k are large-scale matrices having

the following special structures:

$$H_k = D + VEV^T, \quad M_k = \tilde{D} + \tilde{V}\tilde{E}\tilde{V}^T \succ 0,$$

where D, $\tilde{D}$ and $\tilde{E}$ are positive diagonal matrices, V and $\tilde{V}$ have a few columns (say less than 10), and E is a diagonal matrix. LRTR subproblems arise in several contexts. For example, when using TR methods to solve unconstrained or linear-equality constrained minimization problems, the matrix H_k is usually obtained by using a low-rank update (memoryless) formula and the resulting H_k has the structure mentioned above. In such a case, M_k is chosen as either the identity matrix or some other positive definite matrix whose structure is as specified above and depends on the specific problem at hand. We will show that every step of the NE method of [89] can be properly modified to handle the LRTR subproblem and also that the resulting modified NE method is quite efficient and robust for computing NE solutions of large-scale LRTR subproblems.

1.4 Outline and main results of the thesis

In this section, we present the outline and main results of this thesis.

In Chapter 2 we study the limiting behavior of weighted infeasible central paths for semidefinite programming obtained from centrality equations of the form $X^{1/2}SX^{1/2} = \nu W$, where W is a fixed positive definite matrix and $\nu > 0$ is a parameter, under the assumption that the problem has a strictly complementary primal-dual optimal solution. It is shown that a weighted central path as a function of $\sqrt{\nu}$ can be extended analytically beyond 0 and hence that the path converges as $\nu \downarrow 0$. Characterization of the limit points of the path and its normalized first-order derivatives is also provided. As a consequence, it is shown that a weighted central path can have two types of behavior, namely: either it converges as $\Theta(\nu)$ or as $\Theta(\sqrt{\nu})$ depending on whether the matrix W on a certain scaled space is block diagonal or not, respectively. We also derive an error bound on the distance between a point lying in a certain neighborhood of the central path and the set of primal-dual optimal solutions. Finally, in the light of the results of this chapter, we give a characterization of a sufficient condition proposed by Potra and Sheng which guarantees the superlinear convergence of a class of primal-dual interior point SDP algorithms.

In Chapter 3 we first demonstrate that the positive semidefiniteness of a large well-structured sparse symmetric matrix can be represented via the positive semidefiniteness of a collection of smaller matrices linked, in a linear fashion, to the matrix. We derive also the "dual counterpart" of the outlined representation, which expresses the possibility of positive semidefinite completion of a well-structured partially defined symmetric matrix in terms of positive semidefiniteness of a specific collection of fully defined submatrices of the matrix. Using the representations, we then reformulate well-structured large-scale semidefinite problems into smooth convex-concave saddle point problems, which can be solved by a Prox-method developed in [90] with efficiency $\mathcal{O}(\epsilon^{-1})$. Implementations and some numerical results for large-scale Lovász capacity and MAXCUT problems are presented.

In Chapter 4 we develop a long-step primal-dual infeasible path-following algorithm for convex quadratic programming (CQP) whose search directions are computed by means of a preconditioned iterative linear solver. We propose a new linear system, which we refer to as the *augmented normal equation* (ANE), to determine the primal-dual search directions. Since the condition number of the ANE coefficient matrix may become large for degenerate CQP problems, we use a maximum weight basis preconditioner introduced in [99, 111, 124] to precondition this matrix. Using a result obtained in [81], we establish a uniform bound, depending only on the CQP data, for the number of iterations needed by the iterative linear solver to obtain a sufficiently accurate solution to the ANE. Since the iterative linear solver can only generate an approximate solution to the ANE, this solution does not yield a primal-dual search direction satisfying all equations of the primal-dual Newton system. We propose a way to compute an inexact primal-dual search direction so that the equation corresponding to the primal residual is satisfied exactly, while the one corresponding to the dual residual contains a manageable error which allows us to establish a polynomial bound on the number of iterations of our method.

In Chapter 5 we first discuss how the nearly exact (NE) method proposed by Moré and Sorensen [89] for solving trust region subproblems can be modified to solve large-scale "low-rank" trust region TR subproblems efficiently. Our modified algorithm completely avoids

the computation of Cholesky factorizations by instead relying primarily on the Sherman-Morrison formula for computing inverses of "diagonal plus low-rank" type matrices. We also implement a specific version of the modified log-barrier (MLB) algorithm proposed by Polyak [101] where the generated log-barrier subproblems are solved by a trust region method. The corresponding direction finding TR subproblems are of the low-rank type and are then solved by our modified NE method. We finally discuss the computational results of our implementation of the MLB method and its comparison with a version of LANCELOT [23] based on a collection extracted from CUTEr [43] of nonlinear programming problems with simple bound constraints.

CHAPTER II

ERROR BOUNDS AND LIMITING BEHAVIOR OF WEIGHTED PATHS IN SDP

2.1 Preliminary Remarks

Let $\mathcal{S}^n$ denote the space of $n \times n$ real symmetric matrices. We consider the semidefinite programming (SDP) problem

$$(P) \quad \begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & \mathcal{A}X = b, \\ & X \succeq 0, \end{array} \tag{20}$$

and its associated dual SDP problem

$$(D) \quad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \mathcal{A}^* y + S = C, \\ & S \succeq 0, \end{array} \tag{21}$$

where the data consists of $C \in \mathcal{S}^n$, $b \in \Re^m$ and a linear operator $\mathcal{A} : \mathcal{S}^n \to \Re^m$, the primal variable is $X \in \mathcal{S}^n$, and the dual variable consists of $(S, y) \in \mathcal{S}^n \times \Re^m$. For a matrix $V \in \mathcal{S}^n$, the notation $V \succeq 0$ means that V is positive semidefinite. Given a fixed positive definite matrix $W \in \mathcal{S}^n$, $\Delta b \in \Re^m$ and $\Delta C \in \mathcal{S}^n$, our interest in this chapter is to study the set of solutions of the following system of nonlinear equations parametrized by the parameter $\nu > 0$:

$$\mathcal{A}X = b + \nu\Delta b, \quad X \succ 0, \tag{22}$$

$$\mathcal{A}^* y + S = C + \nu\Delta C, \quad S \succ 0, \tag{23}$$

$$X^{1/2} S X^{1/2} = \nu W. \tag{24}$$

Under suitable conditions on $(W, \Delta C, \Delta b)$, it has been shown in Monteiro and Zanjácomo [86] that the above system has a unique solution, denoted by $p(\nu) \equiv (X(\nu), S(\nu), y(\nu))$, for

every $\nu \in (0,1]$. We refer to the path $\nu \in (0,1] \to p(\nu)$ as the $(W, \Delta C, \Delta b)$-weighted central path associated with (P) and (D). The main objective of this chapter is to analyze the limiting behavior of this path as $\nu \downarrow 0$.

When $(W, \Delta C, \Delta b) = (I, 0, 0)$, the path $\nu \in (0,1] \to p(\nu)$ is a part of the central path associated with (P) and (D). Properties of the central path have been extensively studied in several papers due to the important role it plays in the development of interior-points algorithms for cone programming, nonlinear programming and complementarity problems. Early works dealing with the well-definedness, differentiability and limiting behavior of weighted central paths in the context of the linear programming and monotone complementarity problems include [1, 4, 7, 44, 47, 48, 49, 60, 73, 74, 75, 78, 82, 84, 87, 117, 119, 120, 128, 129].

Using the fact that every real algebraic variety has a triangulation, Kojima et al. [59] showed that the central path associated with a monotone linear complementarity problem converges to a solution. In [64], Kojima et al. claims that similar arguments as the ones used in [59] can also be used to show that the central path of a monotone linear semidefinite complementarity problem (which is equivalent to SDP) converges to a solution of the problem. More generally, Drummond and Peterzil [44] established convergence of the central path for analytic convex nonlinear SDP problems. An alternative proof based on a deep result from algebraic geometry (see for example Lemma 3.1 of Milnor [76]) of the convergence of the central path for an SDP problem was given by Halická et al. [52]. Characterization of the limit point of the central path has been obtained by De Klerk et al. [28] and Luo et al. [72] for SDP problems possessing strictly complementary primal-dual optimal solutions. Using an approach based on the implicit function theorem described in Stoer and Wechs [119, 120], Halická [50] showed that the central path of an SDP problem possessing a strictly complementary primal-dual optimal solution can be extended analytically as a function of $\nu > 0$ to $\nu = 0$. For more general SDP problems, the above issues regarding the central path still remain open but some advances have been made on a few papers. These include De Klerk et al. [27] and Goldfarb and Scheinberg [42] who proved that any cluster point of the central path must be a maximally complementary optimal solution. Also, Halická et al.

[51] and Sporre and Forsgren [117] provided partial characterizations of the limit point of the central path as being the analytic center of some convex subset of the optimal solution set and the unique solution of a perturbed log barrier problem over the optimal solution set, respectively. Finally, the recent paper by Cruz Neto et al. [26], which appeared after the release of the first version of the present work, establishes the convergence of the central path for a special class of SDPs which do not satisfy the strict complementarity condition.

Generalization of the notion of weighted central paths from linear programming to SDP problems is a delicate issue. While for a linear programming a weighted central path can be characterized as optimal solutions of certain weighted logarithmic barrier problems, this characterization does not seem to be a good source to obtain a suitable notion of weighted central paths for SDP. Instead, Monteiro and Zanjácomo [86] (see also Monteiro and Pang [83]) work directly with a system consisting of (22), (23) and an equation of the form $\Phi(X,S) = \nu W$, for some suitable map $\Phi : D \subseteq \mathcal{S}^n \times \mathcal{S}^n \to \mathcal{S}^n$, and show that this system has a unique solution for every $\nu \in (0,1]$. Special instances of the map Φ for which the above result applies include the map $(X,S) \to (XS+SX)/2$ and $(X,S) \to X^{1/2}SX^{1/2}$.

Independently to the present work, Preiß and Stoer [107] have proved that the weighted central paths associated with the map $(X,S) \to (XS+SX)/2$ is analytically extendible as functions of $\nu \in (0,1]$ to $\nu = 0$ (see also Lu and Monteiro [70] for another proof of this result). In this chapter, we will be interested only in the second map and its corresponding weighted central paths, i.e., the path of solutions of systems of the form (22)-(24). More specifically, we will investigate the asymptotic properties of the weighted central paths $\nu \in (0,1] \to p(\nu)$ and their derivatives for the special class of SDPs possessing strictly complementary primal-dual optimal solutions. Using a suitable change of variables together with the technique described in [119, 120] based on the implicit function theorem, we prove in Section 2.3 that the path $t \in (0,1] \to p(t^2)$ can be extended analytically to $t = 0$ and we also characterize the limit point of $p(\nu)$ as $\nu \downarrow 0$. In Section 2.4, we characterize the limit of the normalized derivative $p'(\nu)/\|p(\nu)\|$ as $\nu \downarrow 0$. As a consequence, we show that a weighted central path can have two types of behavior, namely: it converges either as $\Theta(\nu)$ or as $\Theta(\sqrt{\nu})$, depending on whether the matrix W on a certain scaled space is block diagonal or not, respectively.

Using these results, we derive in Section 2.5 an error bound on the distance between a point lying in a certain neighborhood of the central path and the set of primal-dual optimal solutions. Finally, we consider in Section 2.6 a sufficient condition proposed by Potra and Sheng [103], which guarantees the superlinear convergence of a large class of primal-dual interior point SDP algorithms, and obtain a characterization of it in terms of the results obtained in this chapter.

The organization of this chapter is as follows. Section 2.2 introduces the assumptions made throughout this chapter and discusses some preliminary known results about weighted central paths. Sections 2.3-2.6 establish the results mentioned in the previous paragraph. Finally, we end this chapter by providing some concluding remarks in Section 2.7.

2.1.1 Notation

The space of symmetric $n \times n$ matrices will be denoted by S^n. Given matrices X and Y in $\Re^{p\times q}$, the standard inner product is defined by $X \bullet Y \equiv \mathrm{Tr}(X^T Y)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. The Euclidean norm and its associated operator norm, i.e., the spectral norm, are both denoted by $\|\cdot\|$. The Frobenius norm of a $p \times q$-matrix X is defined as $\|X\|_F \equiv \sqrt{X \bullet X}$. Given a point f and a set F in a finite dimensional normed vector space, the distance from f to F is defined as $\mathrm{dist}(f,\ F) \equiv \inf_{\tilde{f}\in F} \|f - \tilde{f}\|$. If $X \in \mathcal{S}^n$ is positive semidefinite (resp., definite), we write $X \succeq 0$ (resp., $X \succ 0$). The cone of positive semidefinite (resp., definite) matrices is denoted by $\mathcal{S}^n_+$ (resp., $\mathcal{S}^n_{++}$). Either the identity matrix or operator will be denoted by I. The image (or range) space of a linear operator $\mathcal{A}$ will be denoted by $\mathrm{Im}(\mathcal{A})$; the dimension of the subspace $\mathrm{Im}(\mathcal{A})$, referred to as the rank of $\mathcal{A}$, will be denoted by $\mathrm{rank}(\mathcal{A})$. Given a linear operator $\mathcal{F} : E \to F$ between two finite dimensional inner product spaces $(E, \langle\cdot,\cdot\rangle_E)$ and $(F, \langle\cdot,\cdot\rangle_F)$, its *adjoint* is the unique operator $\mathcal{F}^* : F \to E$ satisfying $\langle \mathcal{F}(u), v\rangle_F = \langle u, \mathcal{F}^*(v)\rangle_E$ for all $u \in E$ and $v \in F$.

Given functions $f : \Omega \to E$ and $g : \Omega \to \Re_{++}$, where Ω is an arbitrary set and E is a normed vector space, and a subset $\tilde{\Omega} \subset \Omega$, we write $f(w) = \mathcal{O}(g(w))$ for all $w \in \tilde{\Omega}$ to mean that there exists $M \geq 0$ such that $\|f(w)\| \leq M g(w)$ for all $w \in \tilde{\Omega}$; moreover, for a function $U : \Omega \to \mathcal{S}^n_{++}$, we write $U(w) = \Theta(g(w))$ for all $w \in \tilde{\Omega}$ if $U(w) = \mathcal{O}(g(w))$ and

$U(w)^{-1} = \mathcal{O}(1/g(w))$ for all $w \in \tilde{\Omega}$. The latter condition is equivalent to the existence of a constant $M > 0$ such that

$$\frac{1}{M} I \preceq \frac{1}{g(w)} U(w) \preceq MI, \quad \forall w \in \Omega.$$

2.2 Preliminaries

In this section, we describe the assumptions that will be used in our presentation. We also describe the weighted central path that will be the subject of our study in this chapter. Some preliminary results about this path are also stated including conditions for its well-definedness.

Throughout this chapter we will be dealing with the pair of dual SDPs (P) and (D) (see (20) and (21), respectively). Denote the feasible sets of (P) and (D) by $\mathcal{F}_P$ and $\mathcal{F}_D$, respectively. Throughout our presentation we make the following assumptions on the pair of problems (P) and (D).

A.1 $\mathcal{A} : \mathcal{S}^n \to \Re^m$ is an onto linear operator;

A.2 There exists a pair of strictly complementary primal-dual optimal solution for (P) and (D), that is a triple $(X^*, S^*, y^*) \in \mathcal{F}_P \times \mathcal{F}_D$ satisfying $X^*S^* = 0$ and $X^* + S^* \succ 0$.

We will assume that Assumptions **A.1** and **A.2** are in force throughout our presentation. Hence, we will state our results without explicitly mentioning them.

Assumption **A.1** is not really crucial for our analysis but it is convenient to ensure that the variables S and y are in one-to-one correspondence. We will see that the dual weighted central path can always be defined in the S-space. The goal of Assumption **A.1** is just to ensure that this path can also be extended to the y-space.

Assumption **A.2** is the one that is commonly used in the analysis of superlinear convergence of interior-point algorithms and it plays an important role in our analysis. In fact, it is a very challenging problem to generalize the analysis of this chapter to the case where Assumption **A.2** is dropped or simply relaxed.

By assumption **A.2**, since $X^*S^* = S^*X^* = 0$, we can diagonalize X^* and S^* simultaneously, i.e. find an orthonormal $P \in \Re^{n \times n}$ such that $P^T X^* P$ and $P^T S^* P$ are both diagonal.

Performing the change of variables $\hat{X} = P^T X P$ and $(\hat{S}, \hat{y}) = (P^T S P, y)$ on problems (P) and (D) yield another pair of primal and dual SDPs which has a primal-dual optimal solution $(\hat{X}^*, \hat{S}^*, \hat{y}^*)$ such that $\hat{X}^*$ and $\hat{S}^*$ are both diagonal. To simplify our notation, we will assume without loss of generality that the original (P) and (D) already have a primal-dual optimal solution (X^*, S^*, y^*) such that

$$X^* = \begin{bmatrix} \Lambda_B & 0 \\ 0 & 0 \end{bmatrix}, \quad S^* = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_N \end{bmatrix}, \tag{25}$$

where $\Lambda_B \equiv \mathrm{diag}(\lambda_1, \cdots, \lambda_K)$, $\Lambda_N \equiv \mathrm{diag}(\lambda_{K+1}, \cdots, \lambda_n)$ for some integer $0 \leq K \leq n$ and some scalars $\lambda_i > 0$, $i = 1, 2, \cdots, n$. Here the subscripts B and N signify the "basic" and "nonbasic" subspaces (following the terminology of linear programming). Throughout this chapter, the decomposition of any $n \times n$ matrix V is always made with respect to the above partition B and N, namely:

$$V = \begin{bmatrix} V_B & V_{BN} \\ V_{NB} & V_N \end{bmatrix},$$

so that V_{BN} and V_{NB} denote the off-diagonal block of V. If $V_{BN} = 0$ and $V_{NB} = 0$, V is called block diagonal, otherwise it is called non-block diagonal.

Notice that $X \in \mathcal{F}_P$ is an optimal solution of (P) if and only if $XS^* = 0$. Hence, by assumption **A.2**, the primal optimal solution set $\mathcal{F}_P^*$ is given by

$$\mathcal{F}_P^* \equiv \{X \in \mathcal{F}_P : \, X_{BN} = 0, \, X_{NB} = 0 \text{ and } X_N = 0\}.$$

Analogously, the dual optimal solution set $\mathcal{F}_D^*$ is given by

$$\mathcal{F}_D^* \equiv \{(S, y) \in \mathcal{F}_D : \, S_{BN} = 0, \; S_{NB} = 0 \text{ and } S_B = 0\}.$$

Define the linear map $\mathcal{G} : \mathcal{S}^n \times \mathcal{S}^n \times \Re^m \to \mathcal{S}^n \times \Re^m$ by

$$\mathcal{G}(X, S, y) \equiv (\mathcal{A}^* y + S - C, \; \mathcal{A} X - b) \tag{26}$$

and the sets $\mathcal{G}_{++}$ and $\mathcal{W}$ by

$$\mathcal{G}_{++} \equiv \mathcal{G}(\mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m), \tag{27}$$

$$\mathcal{W} \equiv \left\{ W \in \mathcal{S}^n_{++} : \, \|W - \nu I\| < \nu/\sqrt{2} \text{ for some } \nu > 0 \right\}. \tag{28}$$

Given $(W, \Delta C, \Delta b) \in \mathcal{S}^n \times \mathcal{S}^n \times \Re^m$, in this chapter we are interested in the solutions of the system of nonlinear equations (22)-(24) parametrized by the parameter $\nu > 0$. The following result gives condition on $(W, \Delta C, \Delta b)$ for system (22)-(24) to have a unique solution for each $\nu \in (0, 1]$.

Proposition 2.2.1 *Assume that $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$. Then, for any $\nu \in (0, 1]$, system (22)-(24) has a unique solution, denoted by $(X(\nu), S(\nu), y(\nu))$. Moreover, the path $\nu \in (0, 1] \to (X(\nu), S(\nu), y(\nu))$ is analytic.*

Proof. By **A.2** and the assumption that $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$, we easily see that $\nu(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$ for all $\nu \in (0, 1]$. The first conclusion of the proposition now follows from Theorem 1(b) of Monteiro and Zanjácomo [86] by letting F, Φ and $\mathcal{V}$ in that theorem be defined as $F = \mathcal{G}$, $\Phi(X, S) = X^{1/2} S X^{1/2}$ for all $(X, S) \in \mathcal{S}^n_+ \times \mathcal{S}^n_+$ and $\mathcal{V} = \mathcal{W}$. The second conclusion follows by applying the analytic version of the implicit function theorem to system (22)-(24) viewed as a function of (X, S, y, ν) and using the fact that the assumption $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$ implies that the Jacobian of this system with respect to (X, S, y) is nonsingular at $(X(\nu), S(\nu), y(\nu), \nu)$ for every $\nu \in (0, 1]$. (See Theorem 2.4 of [85] and the paragraph following it.) ∎

For a given $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$, the path $\nu \in (0, 1] \to (X(\nu), S(\nu), y(\nu))$ will be referred to as the $(W, \Delta C, \Delta b)$-weighted central path. In view of the above proposition, we will assume throughout Sections 2-4 that the following condition is true, without explicitly mentioning it in the statements of the results.

A.3 $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$.

The next result gives some estimates on the size of the blocks of $X(\nu)$ and $S(\nu)$.

Lemma 2.2.2 *For all $\nu > 0$ sufficiently small, we have:*

$$X_B(\nu) = \mathcal{O}(1), \qquad S_N(\nu) = \mathcal{O}(1), \tag{29}$$

$$X_N(\nu) = \mathcal{O}(\nu), \qquad S_B(\nu) = \mathcal{O}(\nu), \tag{30}$$

$$X_{BN}(\nu) = \mathcal{O}(\sqrt{\nu}), \qquad S_{BN}(\nu) = \mathcal{O}(\sqrt{\nu}). \tag{31}$$

Proof. Assume that $\nu > 0$ is sufficiently small and, for notational convenience, let $X \equiv X(\nu)$ and $S \equiv S(\nu)$. Also, let (X^*, S^*, y^*) be as in condition A.2. Since $(\Delta C, \Delta b) \in \mathcal{G}_{++}$, we have $(\Delta C, \Delta b) = \mathcal{G}(X^0, S^0, y^0)$ for some $(X^0, S^0, y^0) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m$. It is easy to see that $A(X - \nu X^0 - (1-\nu)X^*) = 0$ and $S - \nu S^0 - (1-\nu)S^* \in \mathrm{Im}(\mathcal{A}^*)$, and hence that

$$(X - \nu X^0 - (1-\nu)X^*) \bullet (S - \nu S^0 - (1-\nu)S^*) = 0. \tag{32}$$

This equality together with (25) and the fact that $X^* \bullet S^* = 0$, $X \bullet S = \nu \,\mathrm{Tr}(W)$ and all the quantities X, X^0, X^*, S, S^0, S^* are in $\mathcal{S}^n_+$ imply that

$$X \bullet S^0 + X^0 \bullet S \leq \mathrm{Tr}(W) + \xi(\nu) \tag{33}$$

and

$$X_N \bullet S^*_N + X^*_B \bullet S_B = X \bullet S^* + X^* \bullet S \leq \frac{\nu(\mathrm{Tr}(W) + \xi(\nu))}{1-\nu}, \tag{34}$$

where $\xi(\nu) \equiv \nu(X^0 \bullet S^0) + (1-\nu)(X^0 \bullet S^* + X^* \bullet S^0)$. The above two inequalities together with the fact that the matrices $X^0, S^0, X^*_B, S^*_N, X_N, S_B$ are all positive definite clearly imply that (29) and (30) hold. Using the fact that $X(\nu) \succ 0$ and $S(\nu) \succ 0$, we obtain that $X^2_{ij}(\nu) \leq X_{ii}(\nu)X_{jj}(\nu)$ and $S^2_{ij}(\nu) \leq S_{ii}(\nu)S_{jj}(\nu)$ for all i, j. These inequalities together with (29) and (30) imply (31). ∎

The next result gives estimates on the size of the blocks of the matrix $X^{1/2}(\nu) \equiv [X(\nu)]^{1/2}$.

Lemma 2.2.3 *Let $U(\nu) \equiv X^{1/2}(\nu)$ for all $\nu \in (0, 1]$. Then, for all $\nu > 0$ sufficiently small, we have:*

$$U(\nu) = \begin{pmatrix} U_B(\nu) & U_{BN}(\nu) \\ U_{NB}(\nu) & U_N(\nu) \end{pmatrix} = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\nu}) \\ \mathcal{O}(\sqrt{\nu}) & \mathcal{O}(\sqrt{\nu}) \end{pmatrix}.$$

Proof. For notational convenience, let $U = U(\nu)$. Since $X = UU$, we have $X_B = U_B U_B + U_{BN} U^T_{BN}$ and $X_N = U_N U_N + U_{NB} U^T_{NB}$. Hence, we obtain that

$$n\|X_B\| \geq \mathrm{Tr} X_B = \|U_B\|^2_F + \|U_{BN}\|^2_F \geq \max\{\|U_B\|^2, \|U_{BN}\|^2\},$$
$$n\|X_N\| \geq \mathrm{Tr} X_N = \|U_N\|^2_F + \|U_{NB}\|^2_F \geq \max\{\|U_N\|^2, \|U_{NB}\|^2\},$$

from which the result follows in view of (29) and (30). ∎

We end this section by stating a convergence result of the $(W, \Delta C, \Delta b)$-weighted central path to a primal-dual optimal solution of (20) and (21). We do not provide a proof for it since it is similar to the one given in the Appendix of Halická et al. [52].

Proposition 2.2.4 *There exists some $\epsilon > 0$ and an analytic function $\nu : [0, \epsilon) \to (0, 1)$ such that $\nu(0) = 0$ and the path $t \in (0, \epsilon) \to (X(\nu(t)), S(\nu(t)), y(\nu(t)))$ is analytic at $t = 0$. In particular, $(X(\nu(t)), S(\nu(t)), y(\nu(t)))$ converges to some primal-dual optimal solution (X^*, S^*, y^*) as $t \downarrow 0$.*

We observe that Proposition 2.2.4 holds even without requiring Assumption **A.2**. As a consequence, its main advantage is that it holds for any SDP problem. Its main drawbacks are that it neither gives a characterization of the limit point (X^*, S^*, y^*) nor describes how fast $\nu(t)$ converges to 0. These issues and others will be addressed in the remaining sections of this chapter in the context of SDPs satisfying Assumption **A.2**.

2.3 Analyticity of the weighted central path

In the parametrization introduced in the previous section, the weighted central path in general cannot be extended analytically to an interval of the form $(-\epsilon, \infty)$, for some $\epsilon > 0$ (see Corollary 2.4.3). However, in this section we will show that the re-parametrized weighted central path $t \to p(t^2)$ can be extended analytically to an interval as above.

For the sake of brevity, it is convenient to introduce the following definition.

Definition 1 *Let $w : (0, \delta) \to E$ be a given function where $\delta > 0$ and E is a finite dimensional normed vector space. The function w is said to be analytic at 0 if there exist $\epsilon > 0$ and an analytic function $\psi : (-\epsilon, \epsilon) \to E$ such that $w(t) = \psi(t)$ for all $t \in (0, \epsilon)$.*

The basic result that we use to establish that a function $w : (0, \delta) \to E$ is analytic at 0 is the following corollary of the analytic version of the implicit function theorem.

Proposition 2.3.1 *Let $w : (0, \delta) \to E$ be a given function where $\delta > 0$ and E is a finite dimensional normed vector space. Assume that there exists an analytic function $H : \mathcal{O} \times (-\epsilon, \epsilon) \to E$, where $\epsilon > 0$ and $\mathcal{O}$ is an open subset of E, such that $w = w(t)$ is the unique*

solution of $H(w,t) = 0$ *in* $\mathcal{O}$*, for every* $t \in (0,\epsilon)$*. Assume also there exists* $\bar{w} \in \mathcal{O}$ *such that* $H(\bar{w},0) = 0$ *and* $H'_w(\bar{w},0)$ *is nonsingular. Then,*

i) $w = \bar{w}$ *is the unique solution of the system* $H(w,0) = 0$*;*

ii) w *is analytic at* 0 *and, as a consequence,* $\lim_{t\downarrow 0} w(t) = \bar{w}$ *and the limits of all the derivatives of* $w(t)$ *as* $t \downarrow 0$ *exist.*

The following theorem is one of the main results of this section. Its proof will be given at the end of this section.

Theorem 2.3.2 *The re-parametrized* $(W, \Delta C, \Delta b)$*-weighted central path* $t \in (0,1] \rightarrow (X(t^2), S(t^2), y(t^2))$ *is analytic and also analytic at* $t = 0$*. Consequently, the* $(W, \Delta C, \Delta b)$*-weighted central path* $\nu \in (0,1] \rightarrow (X(\nu), S(\nu), y(\nu))$ *converges.*

A key step towards showing the above result is a reformulation of the weighted central path system (22)-(24) as we now discuss. First, observe that (22), (23) and the equations

$$USU = t^2 W, \tag{35}$$

$$UU = X. \tag{36}$$

have $(U, X, S, y) = (U(t^2), X(t^2), S(t^2), y(t^2))$ as its unique solution in $\mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m$, where $U(t^2) \equiv [X(t^2)]^{1/2}$. Letting

$$D_B(t) \equiv \begin{bmatrix} I/t & 0 \\ 0 & I \end{bmatrix}, \qquad D_N(t) \equiv \begin{bmatrix} I & 0 \\ 0 & I/t \end{bmatrix}, \tag{37}$$

and noting that $D_B(t) D_N(t) = I/t$ for every $t \in (0,1]$, we easily see that $U, X, S \in \mathcal{S}^n_{++}$ satisfies (35) and (36) if and only if U, $\tilde{X} \equiv D_N(t) X D_N(t)$ and $\tilde{S} \equiv D_B(t) S D_B(t)$ satisfies

$$[U D_N(t)]\ \tilde{S}\ [D_N(t)\, U] = W, \tag{38}$$

$$[D_N(t)\, U]\ [U D_N(t)] = \tilde{X}. \tag{39}$$

Now, let

$$\mathcal{U}^n = \left\{ U \in \Re^{n\times n} : U_B \in \mathcal{S}^{|B|},\ U_N \in \mathcal{S}^{|N|},\ U_{NB} = 0 \right\},$$

$$\mathcal{U}^n_{++} = \left\{ U \in \mathcal{U}^n : U_B \succ 0,\ U_N \succ 0 \right\}.$$

and define $\mathcal{L} : \mathcal{U}^n \to \Re^{n\times n}$ as

$$\mathcal{L}(U) = \begin{bmatrix} 0 & 0 \\ U_{BN}^T & 0 \end{bmatrix}, \quad \forall\, U \in \mathcal{U}^n.$$

It then follows that $(U, \tilde{X}, \tilde{S})$ satisfies (38) and (39) if and only if $(\tilde{U}, \tilde{X}, \tilde{S})$ with

$$\tilde{U} \equiv \begin{bmatrix} U_B & U_{BN}/t \\ 0 & U_N/t \end{bmatrix},$$

satisfies the equations

$$\left[\tilde{U} + t\,\mathcal{L}(\tilde{U})\right] \tilde{S} \left[\tilde{U} + t\,\mathcal{L}(\tilde{U})\right]^T = W, \tag{40}$$

$$\left[\tilde{U} + t\,\mathcal{L}(\tilde{U})\right]^T \left[\tilde{U} + t\,\mathcal{L}(\tilde{U})\right] = \tilde{X}. \tag{41}$$

Indeed, the above claim follows from the identity

$$UD_N(t) = \begin{bmatrix} U_B & U_{BN}/t \\ U_{NB} & U_N/t \end{bmatrix} = \tilde{U} + t\,\mathcal{L}(\tilde{U}).$$

The above arguments establish the following key result.

Proposition 2.3.3 *Let $(X^*, S^*, y^*) \in \mathcal{F}_P^* \times \mathcal{F}_D^*$ be given. Then, for every $t \in (0, 1]$, the system defined by (40), (41) and the linear equations*

$$\mathcal{A}\left(D_N(t)^{-1}\tilde{X}D_N(t)^{-1} - X^*\right) = t^2\Delta b, \tag{42}$$

$$D_B(t)^{-1}\tilde{S}\,D_B(t)^{-1} - S^* \in t^2\Delta C + \mathrm{Im}(\mathcal{A}^*). \tag{43}$$

has a unique solution, denoted by $(\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$, in $\mathcal{U}_{++}^n \times \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n$. Moreover, the path $t \in (0, 1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ is analytic and, for every $t \in (0, 1]$,

$$\tilde{X}(t) = D_N(t)X(t^2)D_N(t), \quad \tilde{S}(t) = D_B(t)S(t^2)D_B(t), \tag{44}$$

$$\tilde{U}(t) = \begin{bmatrix} U_B(t^2) & U_{BN}(t^2)/t \\ 0 & U_N(t^2)/t \end{bmatrix}. \tag{45}$$

The next result states some basic properties about the accumulation points of the path $t \in (0, 1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ as t approaches 0.

Proposition 2.3.4 *The path* $t \in (0,1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ *remains bounded as* t *approaches* 0 *and any accumulation point* $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ *of this path as* t *approaches* 0 *is in* $\mathcal{U}^n_{++} \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$ *and satisfies the equations*

$$\tilde{U}\tilde{S}\,\tilde{U}^T = W, \tag{46}$$

$$\tilde{U}^T\tilde{U} = \tilde{X}. \tag{47}$$

Proof. By (37) and (44), we have

$$\tilde{X}(t) = \begin{bmatrix} X_B(t^2) & X_{BN}(t^2)/t \\ X_{NB}(t^2)/t & X_N(t^2)/t^2 \end{bmatrix}, \quad \tilde{S}(t) = \begin{bmatrix} S_B(t^2)/t^2 & S_{BN}(t^2)/t \\ S_{NB}(t^2)/t & S_N(t^2) \end{bmatrix}, \tag{48}$$

which, together with Lemma 2.2.2, imply that $(\tilde{X}(t), \tilde{S}(t))$ remains bounded as t approaches 0. Relation (45) and Lemma 2.2.3 imply that $\tilde{U}(t)$ also remains bounded as t approaches 0.

Consider an accumulation point $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ of the path $t \in (0,1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ as t approaches 0. By (44) and (45), we see that $(\tilde{U}(t), \tilde{X}(t), \tilde{S}(t)) \in \mathcal{U}^n_{++} \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$ for all $t \in (0,1]$, and hence we must have $\tilde{U}^* \in \mathcal{U}^n$, $\tilde{X}^* \succeq 0$, $\tilde{S}^* \succeq 0$, $\tilde{U}^*_B \succeq 0$ and $\tilde{U}^*_N \succeq 0$. Thus, to conclude that $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*) \in \mathcal{U}^n_{++} \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$, it suffices to show that $\tilde{U}^*$, $\tilde{X}^*$ and $\tilde{S}^*$ are all invertible. Indeed, since $(\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ satisfies (40) and (41), we conclude upon letting $t \downarrow 0$ that $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ satisfies (46) and (47). This conclusion together with the fact that $W \succ 0$ implies that $\tilde{U}^*$, $\tilde{X}^*$ and $\tilde{S}^*$ are all invertible. ∎

Our next goal is to show that the path $t \in (0,1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t))$ is analytic at $t = 0$. The basic tool we use to establish this fact is the implicit function theorem applied to a specific system of equations parametrized by the parameter $t \in \Re$. A first natural candidate for such a system seems to be the one given by (40), (41), (42) and (43). However, the main drawback of this system is that its Jacobian with respect to $(\tilde{U}, \tilde{X}, \tilde{S})$ is generally singular for $t = 0$ (even though for $t \in (0,1)$ it is always nonsingular). The main cause for this phenomenon is that the "rank" of the linear equations (42) and (43) changes when t becomes 0.

We will now show how the linear equations (42) and (43) can be reformulated into equivalent linear equations for every $t \in (0,1]$. Moreover, the new linear equations have

the property that their rank remains constant for every $t \in \Re$. First note that the linear operator $\mathcal{A} : \mathcal{S}^n \to \Re^m$ can be expressed as

$$\mathcal{A}(X) = \mathcal{A}_B(X_B) + \mathcal{A}_{BN}(X_{BN}) + \mathcal{A}_N(X_N) \equiv (\mathcal{A}_B \ \mathcal{A}_{BN} \ \mathcal{A}_N) \begin{pmatrix} X_B \\ X_{BN} \\ X_N \end{pmatrix}, \qquad (49)$$

for some linear operators $\mathcal{A}_B : \mathcal{S}^{|B|} \to \Re^m$, $\mathcal{A}_{BN} : \Re^{|B|\times|N|} \to \Re^m$ and $\mathcal{A}_N : \mathcal{S}^{|N|} \to \Re^m$.

A well-known result from linear algebra says that any matrix can be put into row-echelon form after a sequence of elementary row operations. A similar type of argument allows one to establish the following result.

Lemma 2.3.5 *Let $\mathcal{A} : \mathcal{S}^n \to \Re^m$ be an onto linear operator. Assume that*

$$i_1 = \operatorname{rank}(\mathcal{A}_B), \quad i_2 = \operatorname{rank}(\mathcal{A}_B \ \mathcal{A}_{BN}) - i_1, \quad i_3 = \operatorname{rank}(\mathcal{A}) - (i_1 + i_2) = m - (i_1 + i_2).$$

Then there exists an isomorphism $T : \Re^m \to \Re^m$ such that $(T \circ \mathcal{A})(X)$ equals

$$\begin{pmatrix} \mathcal{A}_{11}(X_B) & + \mathcal{A}_{12}(X_{BN}) & + \mathcal{A}_{13}(X_N) \\ & \mathcal{A}_{22}(X_{BN}) & + \mathcal{A}_{23}(X_N) \\ & & \mathcal{A}_{33}(X_N) \end{pmatrix} \equiv \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{A}_{13} \\ 0 & \mathcal{A}_{22} & \mathcal{A}_{23} \\ 0 & 0 & \mathcal{A}_{33} \end{pmatrix} \begin{pmatrix} X_B \\ X_{BN} \\ X_N \end{pmatrix},$$

for some linear operators

$$\begin{aligned} &\mathcal{A}_{11} : \mathcal{S}^{|B|} \to \Re^{i_1}, \qquad &&\mathcal{A}_{12} : \Re^{|B|\times|N|} \to \Re^{i_1}, \\ &\mathcal{A}_{13} : \mathcal{S}^{|N|} \to \Re^{i_1}, \qquad &&\mathcal{A}_{22} : \Re^{|B|\times|N|} \to \Re^{i_2}, \\ &\mathcal{A}_{23} : \mathcal{S}^{|N|} \to \Re^{i_2}, \qquad &&\mathcal{A}_{33} : \mathcal{S}^{|N|} \to \Re^{i_3} \end{aligned}$$

such that $\operatorname{rank}(\mathcal{A}_{11}) = i_1,\ \operatorname{rank}(\mathcal{A}_{22}) = i_2,\ \operatorname{rank}(\mathcal{A}_{33}) = i_3.$

We can now reformulate the linear system (42) with the use of Lemma 2.3.5 as follows. Using the fact that

$$D_N^{-1}\tilde{X}D_N^{-1} - X^* = \begin{bmatrix} \tilde{X}_B - X_B^* & t\tilde{X}_{BN} \\ t\tilde{X}_{NB} & t^2\tilde{X}_N \end{bmatrix}$$

and Lemma 2.3.5, we easily see that (42) is equivalent to the linear system

$$\begin{pmatrix} \mathcal{A}_{11} & t\mathcal{A}_{12} & t^2\mathcal{A}_{13} \\ 0 & t\mathcal{A}_{22} & t^2\mathcal{A}_{23} \\ 0 & 0 & t^2\mathcal{A}_{33} \end{pmatrix} \begin{pmatrix} \tilde{X}_B - X_B^* \\ \tilde{X}_{BN} \\ \tilde{X}_N \end{pmatrix} = t^2 \begin{pmatrix} \widetilde{\Delta b_1} \\ \widetilde{\Delta b_2} \\ \widetilde{\Delta b_3} \end{pmatrix}$$

where $\widetilde{\Delta b} \equiv T(\Delta b)$. Dividing the second and third blocks of rows in the above system by t and t^2 respectively, we obtain the following system

$$\begin{pmatrix} \mathcal{A}_{11} & t\mathcal{A}_{12} & t^2\mathcal{A}_{13} \\ 0 & \mathcal{A}_{22} & t\mathcal{A}_{23} \\ 0 & 0 & \mathcal{A}_{33} \end{pmatrix} \begin{pmatrix} \tilde{X}_B - X_B^* \\ \tilde{X}_{BN} \\ \tilde{X}_N \end{pmatrix} = \begin{pmatrix} t^2\widetilde{\Delta b_1} \\ t\widetilde{\Delta b_2} \\ \widetilde{\Delta b_3} \end{pmatrix}. \tag{50}$$

Note that the linear system (50) is equivalent to (42) for every $t \in (0, 1]$. Hence, $\tilde{X}(t)$ satisfies (50) for every $t \in (0, 1]$. A nice feature of (50) is that the operator on its left hand side does not lose full rankness as t becomes 0. We state this fact in the following proposition.

Proposition 2.3.6 *Let $\mathcal{A}_t : \mathcal{S}^n \to \Re^m$ be the operator such that $\mathcal{A}_t(\tilde{X})$ is defined by the left hand side of (50). Then, $t \in \Re \to \mathcal{A}_t$ is a continuous map such* $\operatorname{rank}(\mathcal{A}_t) = m$ *for every $t \in \Re$.*

The linear system (43) can also be reformulated with the aid of Lemma 2.3.5 as follows. First note that by Lemma 2.3.5 we have

$$\operatorname{Im}(\mathcal{A}^*) = \operatorname{Im}\left[(T \circ \mathcal{A})^*\right] = \operatorname{Im}\left[\begin{pmatrix} \mathcal{A}_{11}^* & 0 & 0 \\ \mathcal{A}_{12}^* & \mathcal{A}_{22}^* & 0 \\ \mathcal{A}_{13}^* & \mathcal{A}_{23}^* & \mathcal{A}_{33}^* \end{pmatrix}\right] = \operatorname{Im}\left[\begin{pmatrix} t^2\mathcal{A}_{11}^* & 0 & 0 \\ t^2\mathcal{A}_{12}^* & t\mathcal{A}_{22}^* & 0 \\ t^2\mathcal{A}_{13}^* & t\mathcal{A}_{23}^* & \mathcal{A}_{33}^* \end{pmatrix}\right],$$

for every $t \in (0, 1]$. Hence, for every $t \in (0, 1]$, (43) is equivalent to

$$\begin{pmatrix} t^2\tilde{S}_B \\ t\tilde{S}_{BN} \\ \tilde{S}_N - S_N^* \end{pmatrix} \in t^2 \begin{pmatrix} \Delta C_B \\ \Delta C_{BN} \\ \Delta C_N \end{pmatrix} + \operatorname{Im}\left[\begin{pmatrix} t^2\mathcal{A}_{11}^* & 0 & 0 \\ t^2\mathcal{A}_{12}^* & t\mathcal{A}_{22}^* & 0 \\ t^2\mathcal{A}_{13}^* & t\mathcal{A}_{23}^* & \mathcal{A}_{33}^* \end{pmatrix}\right].$$

Dividing the first and second block of rows in the above system by t^2 and t, respectively, we obtain the system

$$\begin{pmatrix} \tilde{S}_B \\ \tilde{S}_{BN} \\ \tilde{S}_N - S_N^* \end{pmatrix} \in \begin{pmatrix} \Delta C_B \\ t\Delta C_{BN} \\ t^2 \Delta C_N \end{pmatrix} + \mathrm{Im} \left[\begin{pmatrix} \mathcal{A}_{11}^* & 0 & 0 \\ t\mathcal{A}_{12}^* & \mathcal{A}_{22}^* & 0 \\ t^2\mathcal{A}_{13}^* & t\mathcal{A}_{23}^* & \mathcal{A}_{33}^* \end{pmatrix} \right], \tag{51}$$

which is equivalent to (43), and hence satisfied by $\tilde{S}(t)$, for all $t \in (0, 1]$.

Using the definition of $\mathcal{A}_t$ and the fact that $\tilde{X}(t)$ and $\tilde{S}(t)$ satisfy (50) and (51), respectively, for every $t \in (0, 1]$, we conclude that there exists a function $\tilde{y} : (0, 1] \to \Re^m$ such that $(\tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ satisfies

$$\mathcal{A}_t(\tilde{X} - X^*) = \begin{pmatrix} t^2\widetilde{\Delta b_1} \\ t\widetilde{\Delta b_2} \\ \widetilde{\Delta b_3} \end{pmatrix}, \quad \mathcal{A}_t^* \ \tilde{y} + (\tilde{S} - S^*) = \begin{pmatrix} \Delta C_B \\ t\Delta C_{BN} \\ t^2 \Delta C_N \end{pmatrix}, \tag{52}$$

for every $t \in (0, 1]$. Moreover, using Proposition 2.3.6 and the fact that $\{\tilde{S}(t) : t \in (0, 1]\}$ is bounded, we easily see that $\{\tilde{y}(t) : t \in (0, 1]\}$ is also bounded. We have thus established the following result.

Proposition 2.3.7 *There exists a curve $\tilde{y} : \Re_{++} \to \Re^m$ such that $(\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ is the unique solution of (40), (41) and (52) in $\mathcal{U}_{++}^n \times \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n \times \Re^m$ for every $t \in (0, 1]$. Moreover, the path $t \in (0, 1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ remains bounded as t approaches 0 and any of its accumulation points are in $\mathcal{U}_{++}^n \times \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n \times \Re^m$.*

The system formed by (40), (41) and (52) is the one which we will use to establish that the path $t \in (0, 1] \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ is analytic at $t = 0$. This will follow by Proposition 2.3.1 if we can establish that the Jacobian of this system with $t = 0$ with respect to $(\tilde{U}, \tilde{X}, \tilde{S}, \tilde{y})$ is nonsingular as long as $(\tilde{U}, \tilde{X}, \tilde{S})$ is well-centered in the sense that $\|\tilde{U}S\tilde{U} - \nu I\| < \nu/\sqrt{2}$ for some $\nu \in (0, 1]$. The nonsingularity of this Jacobian can be easily seen to be equivalent to showing that $(\widetilde{\Delta U}, \widetilde{\Delta X}, \widetilde{\Delta S}, \widetilde{\Delta y}) = (0, 0, 0, 0) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{S}^n \times \Re^m$

is the only solution of the following linear system:

$$\begin{aligned} \widetilde{\Delta U}\tilde{S}\,\tilde{U}^T + \tilde{U}\widetilde{\Delta S}\,\tilde{U}^T + \tilde{U}\tilde{S}\widetilde{\Delta U}^T &= 0, \\ \widetilde{\Delta U}^T\tilde{U} + \tilde{U}^T\widetilde{\Delta U} &= \widetilde{\Delta X}, \\ \mathcal{A}_0\,\widetilde{\Delta X} &= 0, \\ \mathcal{A}_0^*\,\widetilde{\Delta y} + \widetilde{\Delta S} &= 0. \end{aligned} \tag{53}$$

Before establishing the above fact, we state and prove two technical results.

Lemma 2.3.8 *For any $U \in \mathcal{U}^n_{++}$ and $H \in \mathcal{S}^n$, there exists a unique matrix $V \in \mathcal{U}^n$ such that*

$$H = U^T V + V^T U. \tag{54}$$

Moreover,

$$\|VU^{-1}\|_F \leq \frac{\|U^{-T}HU^{-1}\|_F}{\sqrt{2}}. \tag{55}$$

Proof. The first part of the lemma follows from the fact that the linear map $\Psi_U : \mathcal{U}^n \to \mathcal{S}^n$ defined by $\Psi_U(V) = U^TV + V^TU$ for all $V \in \mathcal{U}^n$ is an isomorphism. Indeed, since its domain and co-domain have the same dimension, it suffices to show that Ψ_U is one-to-one, or equivalently that $U^TV + V^TU = 0$ implies $V = 0$. In turn, this last implication follows from the fact that any solution V of (54) satisfies (55) (simply set $H = 0$ in (55) to conclude that $V = 0$). To show the last claim, we multiply both sides of (54) on the left by U^{-T} and on the right by U^{-1} to obtain

$$U^{-T}HU^{-1} = VU^{-1} + (VU^{-1})^T. \tag{56}$$

Letting $\tilde{U} \equiv VU^{-1}$ and squaring both sides of the above equation, we obtain

$$\|U^{-T}HU^{-1}\|_F^2 = \|\tilde{U} + \tilde{U}^T\|_F^2 = 2\|\tilde{U}\|_F^2 + 2\mathrm{Tr}(\tilde{U}^2). \tag{57}$$

Since

$$\begin{aligned} \tilde{U} = VU^{-1} &= \begin{bmatrix} V_B & V_{BN} \\ 0 & V_N \end{bmatrix} \begin{bmatrix} U_B^{-1} & -U_B^{-1}U_{BN}U_N^{-1} \\ 0 & U_N^{-1} \end{bmatrix} \\ &= \begin{bmatrix} V_BU_B^{-1} & -V_BU_B^{-1}U_{BN}U_N^{-1} + V_{BN}U_N^{-1} \\ 0 & V_NU_N^{-1} \end{bmatrix}, \end{aligned}$$

we have

$$\begin{aligned}\mathrm{Tr}(\tilde{U}^2) &= \mathrm{Tr}((V_B U_B^{-1})^2) + \mathrm{Tr}((V_N U_N^{-1})^2), \\ &= \|U_B^{-1/2} V_B U_B^{-1/2}\|_F^2 + \|U_N^{-1/2} V_N U_N^{-1/2}\|_F^2 \geq 0. \end{aligned} \tag{58}$$

Hence, by (57) and (58), we see that (55) holds. ∎

Lemma 2.3.9 *Suppose that* $\gamma \in [0, 1/\sqrt{2})$ *and that* $(U, S) \in \mathcal{U}^n_{++} \times \mathcal{S}^n$ *is such that* $\|USU^T - \nu I\| \leq \gamma\nu$ *for some* $\nu > 0$. *For some* $H \in \mathcal{S}^n$, *if* $(\Delta U, \Delta X, \Delta S)$ *satisfies*

$$\Delta U S U^T + U \Delta S U^T + U S \Delta U^T = H, \tag{59}$$

$$\Delta U^T U + U^T \Delta U = \Delta X, \tag{60}$$

$$\Delta X \bullet \Delta S = 0, \tag{61}$$

then

$$\max\{\nu\|U^{-T}\Delta X U^{-1}\|_F, \|U\Delta S U^T\|_F\} \leq \frac{\|H\|_F}{(1 - \sqrt{2}\gamma)}. \tag{62}$$

Proof. Multiplying both sides of (60) on the left by U^{-T} and on the right by U^{-1} to obtain

$$U^{-T}\Delta U^T + \Delta U U^{-1} = U^{-T}\Delta X U^{-1}.$$

By this equality and (59), we have

$$\nu U^{-T}\Delta X U^{-1} + U\Delta S U^T = H - \Delta U U^{-1}(USU^T - \nu I) - (USU^T - \nu I)U^{-T}\Delta U^T. \tag{63}$$

Taking the Frobenius norm on both sides of this equality and using (55) and (61), we obtain

$$\begin{aligned} \max \quad & \{\nu\|U^{-T}\Delta X U^{-1}\|_F, \|U\Delta S U^T\|_F\} \\ \leq \quad & \left(\nu^2\|U^{-T}\Delta X U^{-1}\|_F^2 + \|U\Delta S U^T\|_F^2\right)^{1/2} \\ = \quad & \left\|H - \Delta U U^{-1}(USU^T - \nu I) - (USU^T - \nu I)U^{-T}\Delta U^T\right\|_F \\ \leq \quad & \|H\|_F + 2\|\Delta U U^{-1}\|_F\|USU^T - \nu I\| \\ \leq \quad & \|H\|_F + \sqrt{2}\gamma\nu\|U^{-T}\Delta X U^{-1}\|_F, \end{aligned} \tag{64}$$

which clearly implies that

$$\nu\|U^{-T}\Delta X U^{-1}\|_F \leq \frac{\|H\|_F}{(1-\sqrt{2}\gamma)}. \tag{65}$$

Using this last inequality to bound the right hand side of (64), we obtain (62). ∎

As an immediate consequence of the above lemma, we obtain the following corollary.

Corollary 2.3.10 *Assume that* $(\tilde{U},\tilde{S}) \in \mathcal{U}^n_{++}\times\mathcal{S}^n$ *is such that* $\|USU^T-\nu I\| < \nu/\sqrt{2}$ *for some* $\nu>0$. *Then, system (53) has* $(\widetilde{\Delta U},\widetilde{\Delta X},\widetilde{\Delta S},\widetilde{\Delta y}) = (0,0,0,0)$ *as its unique solution.*

Proof. The last two equations of system (53) imply that $\widetilde{\Delta X}\bullet\widetilde{\Delta S}=0$. Using this identity and the first two equations of (53), by Lemma 2.3.9 we easily obtain that $\widetilde{\Delta X}=0$ and $\widetilde{\Delta S}=0$, which together with the second equation of (53) and Lemma 2.3.8 implies $\widetilde{\Delta U}=0$. Also, $\widetilde{\Delta y}=0$ follows from the fact that $\mathcal{A}^*$ is one-to-one and the last equation of (53). ∎

We are ready to establish the analyticity of the path $t\in(0,1]\to(\tilde{U}(t),\tilde{X}(t),\tilde{y}(t),\tilde{S}(t))$.

Theorem 2.3.11 *Let* $(X^*,S^*,y^*)\in\mathcal{F}_P^*\times\mathcal{F}_D^*$ *be given. There hold:*

i) the path $t\in(0,1]\to\tilde{p}(t)\equiv(\tilde{U}(t),\tilde{X}(t),\tilde{S}(t),\tilde{y}(t))$, *where* $(\tilde{U}(t),\tilde{X}(t),\tilde{S}(t),\tilde{y}(t))$ *is the unique solution of (40), (41), (52) in* $\mathcal{U}^n_{++}\times\mathcal{S}^n_{++}\times\mathcal{S}^n_{++}\times\Re^m$, *is analytic and also analytic at* 0; *consequently,* $\tilde{p}(t)$ *and all its k-th order derivatives,* $k\geq 1$, *converge as* $t\downarrow 0$;

ii) $(\tilde{U}^*,\tilde{X}^*,\tilde{S}^*,\tilde{y}^*)\equiv\lim_{t\downarrow 0}(\tilde{U}(t),\tilde{X}(t),\tilde{S}(t),\tilde{y}(t))$ *is the unique solution of the system defined by (46), (47) and*

$$\mathcal{A}_0(\tilde{X}-X^*)=\begin{pmatrix}0\\0\\\widetilde{\Delta b_3}\end{pmatrix},\quad \mathcal{A}_0^*\,\tilde{y}+(\tilde{S}-S^*)=\begin{pmatrix}\Delta C_B\\0\\0\end{pmatrix}; \tag{66}$$

iii) $(\widetilde{\delta U}^*,\widetilde{\delta X}^*,\widetilde{\delta S}^*,\widetilde{\delta y}^*)\equiv\lim_{t\downarrow 0}(\dot{\tilde{U}}(t),\dot{\tilde{X}}(t),\dot{\tilde{S}}(t),\dot{\tilde{y}}(t))$ *is the unique solution of the linear system defined by*

$$\widetilde{\delta U}\,\tilde{S}^*(\tilde{U}^*)^T+\tilde{U}^*\tilde{S}^*\,\widetilde{\delta U}^T+\tilde{U}^*\widetilde{\delta S}\,(\tilde{U}^*)^T=-\left[\mathcal{L}(\tilde{U}^*)\tilde{S}^*(\tilde{U}^*)^T+\tilde{U}^*\tilde{S}^*\mathcal{L}(\tilde{U}^*)^T\right], \tag{67}$$

$$\widetilde{\delta U}^T \tilde{U}^* + (\tilde{U}^*)^T \widetilde{\delta U} - \widetilde{\delta X} = -\left[\mathcal{L}(\tilde{U}^*)^T \tilde{U}^* + (\tilde{U}^*)^T \mathcal{L}(\tilde{U}^*)\right], \tag{68}$$

$$\mathcal{A}_0 \widetilde{\delta X} = -\mathcal{B}_0 \tilde{X}^* + \begin{pmatrix} 0 \\ \widetilde{\Delta b_2} \\ 0 \end{pmatrix}, \quad \mathcal{A}_0^* \widetilde{\delta y} + \widetilde{\delta S} = -\mathcal{B}_0^* \tilde{y}^* + \begin{pmatrix} 0 \\ \Delta C_{BN} \\ 0 \end{pmatrix}, \tag{69}$$

where

$$\mathcal{B}_0 \equiv \begin{pmatrix} 0 & \mathcal{A}_{12} & 0 \\ 0 & 0 & \mathcal{A}_{23} \\ 0 & 0 & 0 \end{pmatrix}.$$

Proof. The proof of theorem is based on Proposition 2.3.1. Indeed, let $E = \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{S}^n \times \Re^m$, $\mathcal{O} = \mathcal{U}^n_{++} \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m$, $\delta = \epsilon = 1$, $w : (0,1) \to E$ denote the path $t \in (0,1) \to (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ and $H(w,t) = H(\tilde{U}, \tilde{X}, \tilde{S}, \tilde{y}, t)$ be the map determined by system (40), (41), (52). By Proposition 2.3.7, the path $\tilde{p}(t) = (\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ has an accumulation point $w^* = (\tilde{U}^*, \tilde{X}^*, \tilde{S}^*, \tilde{y}^*)$ in $\mathcal{O}$ and, by Corollary 2.3.10, it follows that $H'_w(w^*, 0)$ is nonsingular since (46) with $(\tilde{U}, \tilde{X}, \tilde{S}) = (\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ implies that $\|\tilde{U}^* \tilde{S}^* (\tilde{U}^*)^T - \nu I\| = \|W - \nu I\| < \nu/\sqrt{2}$. Hence, i) and ii) follow directly from Proposition 2.3.1. Differentiating the identity $H(\tilde{p}(t), t) = 0$ with respect to t and letting $t \downarrow 0$, we conclude that $\delta w = \delta w^* \equiv (\widetilde{\delta U}^*, \widetilde{\delta X}^*, \widetilde{\delta S}^*, \widetilde{\delta y}^*)$ satisfies

$$H'_w(w^*, 0)\delta w = -H'_t(w^*, 0).$$

Statement iii) now follows from the fact that $H'_w(w^*, 0)$ is nonsingular and the latter system is equivalent to (67)-(69). ∎

The proof of Theorem 2.3.2 is now obvious. Indeed, the analyticity of the map $t \to (X(t^2), S(t^2))$ follows from (48) and the analyticity of $t \to (\tilde{X}(t), \tilde{S}(t))$. The analyticity of $t \to y(t^2)$ follows from the analyticity of $t \to S(t^2)$ and Assumption A.1. The last statement of the theorem is obvious.

In the remainder of this chapter, we will let $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*, \tilde{y}^*)$ and $(\widetilde{\delta U}^*, \widetilde{\delta X}^*, \widetilde{\delta S}^*, \widetilde{\delta y}^*)$ denote the limits of $(\tilde{U}(t), \tilde{X}(t), \tilde{S}(t), \tilde{y}(t))$ and $(\dot{\tilde{U}}(t), \dot{\tilde{X}}(t), \dot{\tilde{S}}(t), \dot{\tilde{y}}(t))$, respectively, as $t \downarrow 0$ (as in Theorem 2.3.11 above). Observe that Theorem 2.3.11 provides a characterization of

$(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*, \tilde{y}^*)$ as being the unique solution of a certain system of equations which arises by first performing some transformations to the original weighted central path system, and then setting $t = 0$ in the resulting system. Hence, it is reasonable to expect that the linear equations (66) can be entirely described in terms of the original data $(W, \mathcal{A}, C, \Delta C, b, \Delta b)$. Indeed, the following result gives this alternative description of (66).

Theorem 2.3.12 *$(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ is the unique solution of the system given by (46), (47) and the linear equations*

$$\mathcal{A}_B(\tilde{X}_B) = b, \quad \mathcal{A}_{BN}(\tilde{X}_{BN}) \in \mathrm{Im}(\mathcal{A}_B), \quad \mathcal{A}_N(\tilde{X}_N) \in \Delta b + \mathrm{Im}(\mathcal{A}_B\ \mathcal{A}_{BN}), \tag{70}$$

$$\tilde{S}_B \in \Delta C_B + \mathrm{Im}(\mathcal{A}_B^*), \begin{pmatrix} 0 \\ \tilde{S}_{BN} \end{pmatrix} \in \mathrm{Im} \begin{pmatrix} \mathcal{A}_B^* \\ \mathcal{A}_{BN}^* \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \tilde{S}_N \end{pmatrix} \in C + \mathrm{Im} \begin{pmatrix} \mathcal{A}_B^* \\ \mathcal{A}_{BN}^* \\ \mathcal{A}_N^* \end{pmatrix}. \tag{71}$$

Proof. From Theorem 2.3.11(ii), it suffices to show that (66) is equivalent to (70) and (71). Since the first equation of (66) is the same as (50) with $t = 0$, we have that the first equation of (66) holds if and only if

$$\mathcal{A}_{11}(\tilde{X}_B) = \mathcal{A}_{11}(X_B^*), \quad \mathcal{A}_{22}(\tilde{X}_{BN}) = 0, \quad \mathcal{A}_{33}(\tilde{X}_N) = \widetilde{\Delta b_3}. \tag{72}$$

By Lemma 2.3.5, the first identity in (72) can be written as

$$(T \circ \mathcal{A}) \begin{pmatrix} \tilde{X}_B \\ 0 \\ 0 \end{pmatrix} = (T \circ \mathcal{A}) \begin{pmatrix} X_B^* \\ 0 \\ 0 \end{pmatrix},$$

and hence it is equivalent to $\mathcal{A}_B(\tilde{X}_B) = \mathcal{A}_B(X_B^*) = b$, in view of relation (49) and the fact that T is an isomorphism. By Lemma 2.3.5 and the fact that $\mathcal{A}_{11}$ is onto, the second identity in (72) holds if and only if

$$(T \circ \mathcal{A}) \begin{pmatrix} \hat{X}_B \\ \tilde{X}_{BN} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

for some $\hat{X}_B \in \mathcal{S}^{|B|}$, and hence it is equivalent to $\mathcal{A}_{BN}(\tilde{X}_{BN}) \in \mathrm{Im}(\mathcal{A}_B)$, in view of (49) and the fact that T is an isomorphism. Using Lemma 2.3.5 again and the fact that $\mathcal{A}_{11}$ and $\mathcal{A}_{22}$ are onto, we easily see that the last identity in (72) holds if and only if

$$(T \circ \mathcal{A}) \begin{pmatrix} \check{X}_B \\ \check{X}_{BN} \\ \tilde{X}_N \end{pmatrix} = \begin{pmatrix} \widetilde{\Delta b_1} \\ \widetilde{\Delta b_2} \\ \widetilde{\Delta b_3} \end{pmatrix} = T(\Delta b)$$

for some $(\check{X}_B, \check{X}_{BN}) \in \mathcal{S}^{|B|} \times \Re^{|B| \times |N|}$, and hence it is equivalent to $\mathcal{A}_N(\tilde{X}_N) \in \Delta b + \mathrm{Im}(\mathcal{A}_B \ \mathcal{A}_{BN})$, in view of (49) and the fact that T is an isomorphism. We have thus shown that the first equation of (66) is equivalent to (70).

The fact that the second equation of (66) holds if and only if (71) holds can be proved in a similar way as above. ∎

The following result gives an alternative characterization of $(\widetilde{\delta U}^*, \widetilde{\delta X}^*, \widetilde{\delta S}^*)$ involving the original data $(W, \mathcal{A}, C, \Delta C, b, \Delta b)$.

Theorem 2.3.13 *$(\widetilde{\delta U}^*, \widetilde{\delta X}^*, \widetilde{\delta S}^*)$ is the unique solution of the linear system of equations (67), (68) and*

$$\begin{bmatrix} \mathcal{A}_B & \mathcal{A}_{BN} \end{bmatrix} \begin{bmatrix} \widetilde{\delta X_B} \\ \tilde{X}^*_{BN} \end{bmatrix} = 0, \quad \begin{bmatrix} \mathcal{A}_{BN} & \mathcal{A}_N \end{bmatrix} \begin{bmatrix} \widetilde{\delta X}_{BN} \\ \tilde{X}^*_N \end{bmatrix} \in \Delta b + \mathrm{Im}(\mathcal{A}_B),$$

$$\mathcal{A}_N(\widetilde{\delta X}_N) \in \mathrm{Im}(\mathcal{A}_B \ \mathcal{A}_{BN}), \tag{73}$$

$$\widetilde{\delta S}_B \in \mathrm{Im}(\mathcal{A}^*_B), \quad \begin{pmatrix} \tilde{S}^*_B \\ \widetilde{\delta S}_{BN} \end{pmatrix} \in \begin{pmatrix} \Delta C_B \\ \Delta C_{BN} \end{pmatrix} + \mathrm{Im} \left[\begin{pmatrix} \mathcal{A}^*_B \\ \mathcal{A}^*_{BN} \end{pmatrix} \right],$$

$$\begin{pmatrix} 0 \\ \tilde{S}^*_{BN} \\ \widetilde{\delta S}_N \end{pmatrix} \in \mathrm{Im} \left[\begin{pmatrix} \mathcal{A}^*_B \\ \mathcal{A}^*_{BN} \\ \mathcal{A}^*_N \end{pmatrix} \right]. \tag{74}$$

Proof. From Theorem 2.3.11(iii), it suffices to show that (69) is equivalent to (73)

and (74). Observe that the first equation of (69) can be written as

$$\begin{aligned} \mathcal{A}_{11}(\widetilde{\delta X_B}) \;+\; \mathcal{A}_{12}(\tilde{X}^*_{BN}) &= 0, \\ \mathcal{A}_{22}(\widetilde{\delta X_{BN}}) \;+\; \mathcal{A}_{23}(\tilde{X}^*_{N}) &= \widetilde{\Delta b_2}, \\ \mathcal{A}_{33}(\widetilde{\delta X_N}) &= 0. \end{aligned} \tag{75}$$

Using Lemma 2.3.5, the fact that $\mathcal{A}_{11}$ and $\mathcal{A}_{22}$ are onto and the identities $\mathcal{A}_{22}\tilde{X}^*_{BN} = 0$ and $\mathcal{A}_{33}\tilde{X}^*_N = \widetilde{\Delta b_3}$ which hold in view of Theorem 2.3.11(ii), we easily see that the above three equations are respectively equivalent to

$$(T \circ \mathcal{A}) \begin{pmatrix} \widetilde{\delta X_B} \\ \tilde{X}^*_{BN} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (T \circ \mathcal{A}) \begin{pmatrix} \hat{X}_B \\ \widetilde{\delta X_{BN}} \\ \tilde{X}^*_N \end{pmatrix} = \begin{pmatrix} \widetilde{\Delta b_1} \\ \widetilde{\Delta b_2} \\ \widetilde{\Delta b_3} \end{pmatrix},$$

$$(T \circ \mathcal{A}) \begin{pmatrix} \check{X}_B \\ \check{X}_{BN} \\ \widetilde{\delta X_N} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

for some $\hat{X}_B, \check{X}_B \in \mathcal{S}^{|B|}$ and $\check{X}_{BN} \in \Re^{|B|\times|N|}$. The latter conditions in turn are equivalent to (73) in view of (49) and the facts that $\widetilde{\Delta b} = T(\Delta b)$ and T is an isomorphism.

Using similar arguments as to ones used above, it can be shown that the second equation of (69) holds if and only if (74) holds. ∎

2.4 Limiting behavior of the derivative of the weighted central path

In this section, we will first characterize the limit of the normalized derivatives of a weighted central path as ν approaches 0. We then show that a weighted central path can have two types of behavior, namely: either it converges as $\Theta(\nu)$ or as $\Theta(\sqrt{\nu})$ depending on whether the matrix W on a certain scaled space is block diagonal or not, respectively.

Theorem 2.4.1 $\lim_{\nu\downarrow 0} \sqrt{\nu}\,(\dot{X}(\nu), \dot{S}(\nu), \dot{y}(\nu))$ *exists and satisfies*

$$\lim_{\nu\downarrow 0} \sqrt{\nu}\,\dot{X}(\nu) = \begin{bmatrix} \widetilde{\delta X}^*_B/2 & \tilde{X}^*_{BN}/2 \\ \tilde{X}^*_{NB}/2 & 0 \end{bmatrix}, \quad \lim_{\nu\downarrow 0} \sqrt{\nu}\,\dot{S}(\nu) = \begin{bmatrix} 0 & \tilde{S}^*_{BN}/2 \\ \tilde{S}^*_{NB}/2 & \widetilde{\delta S}^*_N/2 \end{bmatrix}. \tag{76}$$

Proof. By (48), we have

$$X(t^2) = \begin{bmatrix} \tilde{X}_B(t) & t\tilde{X}_{BN}(t) \\ t\tilde{X}_{NB}(t) & t^2\tilde{X}_N(t) \end{bmatrix}, \quad S(t^2) = \begin{bmatrix} t^2\tilde{S}_B(t) & t\tilde{S}_{BN}(t) \\ t\tilde{S}_{NB}(t) & \tilde{S}_N(t) \end{bmatrix} \tag{77}$$

Differentiating both sides with respect to t, letting $t \downarrow 0$, and using Theorem 2.3.11, we obtain (76) upon letting $\nu = t^2$. ∎

We establish one technical lemma as follows, which gives a characterization of block diagonal weighted matrix W. This lemma will play a crucial role in further analyzing the limiting behavior of derivatives of the weighted central path.

Lemma 2.4.2 *The following statements hold:*

i) $\tilde{X}^*_{BN} \bullet \tilde{S}^*_{BN} = 0$;

ii) $\tilde{X}^*_{BN} = \tilde{S}^*_{BN} = 0$ *if and only if* $W_{BN} = 0$.

Proof. Statement i) follows from the fact that $\tilde{X}^*_{BN}$ and $\tilde{S}^*_{BN}$ satisfy the second equations in (70) and (71), respectively, which can be easily seen to determine two orthogonal complementary subspaces in $\Re^{|B|\times|N|}$.

We now show ii). Using the fact that $(\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ satisfies (46) and (47), it is easy to see that

$$W_{BN} = \tilde{U}^*_B\tilde{S}^*_{BN}\tilde{U}^*_N + \tilde{U}^*_{BN}\tilde{S}^*_N\tilde{U}^*_N, \quad \tilde{X}^*_B = (\tilde{U}^*_B)^2, \quad \tilde{X}^*_{BN} = \tilde{U}^*_B\tilde{U}^*_{BN}. \tag{78}$$

By Proposition 2.3.4, we know that $(\tilde{X}^*, \tilde{S}^*, \tilde{U}^*) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times U^n_{++}$, and hence $\tilde{X}^*_B \succ 0$, $\tilde{S}^*_N \succ 0$, $\tilde{U}^*_B \succ 0$, $\tilde{U}^*_N \succ 0$. Thus, the last relation in (78) implies that

$$\tilde{X}^*_{BN} = 0 \iff \tilde{U}^*_{BN} = 0. \tag{79}$$

Assume first that $\tilde{X}^*_{BN} = \tilde{S}^*_{BN} = 0$. Then, (79) and the first equation in (78) immediately imply that $W_{BN} = 0$. Assume now that $W_{BN} = 0$. Using (78), we obtain

$$\begin{aligned} \tilde{X}^*_B\tilde{S}^*_{BN} + \tilde{X}^*_{BN}\tilde{S}^*_N &= (\tilde{U}^*_B)^2\tilde{S}^*_{BN} + \tilde{U}^*_B\tilde{U}^*_{BN}\tilde{S}^*_N \\ &= \tilde{U}^*_B\left(\tilde{U}^*_B\tilde{S}^*_{BN}\tilde{U}^*_N + \tilde{U}^*_{BN}\tilde{S}^*_N\tilde{U}^*_N\right)(\tilde{U}^*_N)^{-1} \\ &= \tilde{U}^*_B W_{BN}(\tilde{U}^*_N)^{-1} = 0. \end{aligned}$$

Multiplying the above equation on the left by $(\tilde{X}_B^*)^{-1/2}$ and on the right by $(\tilde{S}_N^*)^{-1/2}$, squaring both sides of the resulting expression and using i), we conclude that

$$\|(\tilde{X}_B^*)^{1/2}\tilde{S}_{BN}^*(\tilde{S}_N^*)^{-1/2}\|_F^2 + \|(\tilde{X}_B^*)^{-1/2}\tilde{X}_{BN}^*(\tilde{S}_N^*)^{1/2}\|_F^2 = 0,$$

from which it follows that $\tilde{X}_{BN}^* = \tilde{S}_{BN}^* = 0$. ∎

From Lemma 2.4.2 and Theorem 2.4.1, the following corollary follows.

Corollary 2.4.3 *If $W_{BN} \neq 0$ then at least one of the limits in (76) is nonzero and*

$$\|(X(\nu), S(\nu), y(\nu)) - (X^*, S^*, y^*)\| = \Theta(\sqrt{\nu}).$$

Proof. Assume that $W_{BN} \neq 0$. By Lemma 2.4.2(ii), we have that either $\tilde{X}_{BN}^* \neq 0$ or $\tilde{S}_{BN}^* \neq 0$, which together with (76) implies the first claim of the corollary. The second claim follows directly from (76) and the equality

$$\lim_{\nu \downarrow 0} \frac{(X(\nu), S(\nu), y(\nu)) - (X^*, S^*, y^*)}{\sqrt{\nu}} = \lim_{\nu \downarrow 0} 2\sqrt{\nu} \left(\dot{X}(\nu), \dot{S}(\nu), \dot{y}(\nu)\right),$$

which holds due to Theorem 2.3.2. ∎

From Corollary 2.4.3, we immediately see that the weighted central path as a function of ν in general cannot be extended analytically to an interval of the form $(-\epsilon, \infty)$, for some $\epsilon > 0$. Theorem 2.4.1 and Corollary 2.4.3 give a precise characterization of how the primal-dual weighted central path approaches its limit (X^*, S^*, y^*) for the case when W is non-block diagonal, that is $W_{BN} \neq 0$. However, it is still possible for one of the limits in (76) to be equal to zero in this situation. The following result claims that in this case the corresponding primal or dual weighted central path converges towards (X^*, S^*, y^*) at a $\Theta(\nu)$ rate of convergence.

Theorem 2.4.4 *The following statements hold:*

i) If $\lim_{\nu \downarrow 0} \sqrt{\nu}\, \dot{X}(\nu) = 0$ then $X(\nu) - X^* = \Theta(\nu)$ and

$$\lim_{\nu \downarrow 0} \dot{X}(\nu) = \begin{bmatrix} \widetilde{\delta^{(2)} X}_B^*/2 & \widetilde{\delta X}_{BN}^* \\ (\widetilde{\delta X}_{BN}^*)^T & \tilde{X}_N^* \end{bmatrix}, \tag{80}$$

where $\widetilde{\delta^{(2)} X}_B^* \equiv \lim_{t \downarrow 0} \ddot{\tilde{X}}_B(t)$;

ii) If $\lim_{\nu \downarrow 0} \sqrt{\nu}\, \dot{S}(\nu) = 0$ *then* $\|(S(\nu), y(\nu)) - (S^*, y^*)\| = \Theta(\nu)$ *and*

$$\lim_{\nu \downarrow 0} \dot{S}(\nu) = \begin{bmatrix} \widetilde{\delta^{(2)} S}_B^*/2 & \widetilde{\delta S}_{BN}^* \\ (\widetilde{\delta S}_{BN}^*)^T & \tilde{S}_N^* \end{bmatrix}, \tag{81}$$

where $\widetilde{\delta^{(2)} S}_B^* \equiv \lim_{t \downarrow 0} \ddot{\tilde{S}}_B(t)$.

Proof. To prove i), assume that $\lim_{\nu \downarrow 0} \sqrt{\nu}\, \dot{X}(\nu) = 0$. By Theorem 2.4.1, we must have $\widetilde{\delta X}_B^* = 0$ and $\tilde{X}_{BN}^* = 0$. Differentiating both sides of the first identity in (77) with respect to t and then dividing the resulting identity by $2t$, we obtain that

$$\dot{X}(t^2) = \begin{bmatrix} \dot{\tilde{X}}_B(t)/(2t) & \tilde{X}_{BN}(t)/(2t) + \dot{\tilde{X}}_{BN}(t)/2 \\ \tilde{X}_{BN}(t)^T/(2t) + \dot{\tilde{X}}_{BN}(t)^T/2 & \tilde{X}_N(t) + t\dot{\tilde{X}}_N(t)/2 \end{bmatrix}.$$

Using the fact that $\widetilde{\delta X}_B^* = 0$ and $\tilde{X}_{BN}^* = 0$ and and using Theorem 2.3.11, we obtain (80) upon letting $\nu = t^2 \downarrow 0$. The conclusion that $X(\nu) - X^* = \Theta(\nu)$ follows immediately from (80) and the fact $\tilde{X}_N^* \succ 0$. Using the same arguments as above and assumption **A.1**, we can similarly show ii). ∎

The remainder of this section considers the case when W is block diagonal, that is $W_{BN} = 0$. We will show in this case that two limits in (76) are equal to zero, and hence that $\lim_{\nu \downarrow 0}(\dot{X}(\nu), \dot{S}(\nu), \dot{y}(\nu))$ exists and $\|(X(\nu), S(\nu), y(\nu)) - (X^*, S^*, y^*)\| = \Theta(\nu)$ due to Theorem 2.4.4.

Note that to establish the above claim, it suffices to show that $\widetilde{\delta X}_B^* = 0$ and $\widetilde{\delta S}_N^* = 0$ in view of Lemma 2.4.2(ii). Before showing this fact, we state two technical results from Monteiro and Tsuchiya [85].

Lemma 2.4.5 (Lemma 2.1 of [85]) *For every* $A \in \mathcal{S}_{++}^n$ *and* $H \in \mathcal{S}^n$*, the equation* $AU + UA = H$ *has a unique solution* $U \in \mathcal{S}^n$*. Moreover, this solution satisfies* $\|AU\|_F \leq \|H\|_F/\sqrt{2}$.

Lemma 2.4.6 (Lemma 2.3 of [85]) *Suppose that* $\gamma \in [0, 1/\sqrt{2})$ *and that* $(U, S) \in \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n$ *is such that* $\|USU - \nu I\| \leq \gamma\nu$ *for some* $\nu > 0$*. If* $(\Delta X, \Delta S, \Delta U) \in \mathcal{S}^n \times \mathcal{S}^n \times \mathcal{S}^n$ *is*

a solution of the following system

$$\begin{aligned} \Delta USU + US\Delta U + U\Delta SU &= H, \\ \Delta UU + U\Delta U &= \Delta X, \\ \Delta X \bullet \Delta S &= 0 \end{aligned}$$

and $H \in \mathcal{S}^n$*, then*

$$\max\left\{\nu\|U^{-1}\Delta X U^{-1}\|_F, \|U\Delta S U\|_F\right\} \leq \frac{\|H\|_F}{(1-\sqrt{2}\gamma)}.$$

We are now ready to show that $\widetilde{\delta X}^*_B = 0$ and $\widetilde{\delta S}^*_N = 0$.

Lemma 2.4.7 *If* $W_{BN} = 0$ *then*

$$\begin{aligned} \widetilde{\delta X}^*_B = \widetilde{\delta S}^*_B = \widetilde{\delta U}^*_B = 0, \\ \widetilde{\delta X}^*_N = \widetilde{\delta S}^*_N = \widetilde{\delta U}^*_N = 0. \end{aligned}$$

Proof. From Lemma 2.4.2(ii), we know that $\tilde{X}^*_{BN} = \tilde{S}^*_{BN} = 0$. Using this identity and the fact that $(\widetilde{\delta X}^*_B, \widetilde{\delta S}^*_B)$ and $(\widetilde{\delta X}^*_N, \widetilde{\delta S}^*_N)$ satisfy the first and third equations of (73) and (74), respectively, we obtain that

$$\widetilde{\delta X}^*_B \bullet \widetilde{\delta S}^*_B = 0, \quad \widetilde{\delta X}^*_N \bullet \widetilde{\delta S}^*_N = 0. \tag{82}$$

By (79), $\tilde{X}^*_{BN} = 0$ implies $\tilde{U}^*_{BN} = 0$, and thus $\mathcal{L}(\tilde{U}^*) = 0$. Hence, by (67) and (68), we have

$$\begin{aligned} \widetilde{\delta U}^* \tilde{S}^* (\tilde{U}^*)^T + \tilde{U}^* \tilde{S}^* (\widetilde{\delta U}^*)^T + \tilde{U}^* \widetilde{\delta S}^* (\tilde{U}^*)^T &= 0, \\ (\widetilde{\delta U}^*)^T \tilde{U}^* + (\tilde{U}^*)^T \widetilde{\delta U}^* &= \widetilde{\delta X}^*. \end{aligned}$$

These equations together with the fact that $\tilde{S}^*_{BN} = 0$ and $\tilde{U}^*_{BN} = 0$ can be easily seen to imply that

$$\widetilde{\delta U}^*_B \, \tilde{S}^*_B \, \tilde{U}^*_B + \tilde{U}^*_B \, \tilde{S}^*_B \, \widetilde{\delta U}^*_B + \tilde{U}^*_B \, \widetilde{\delta S}^*_B \, \tilde{U}^*_B = 0, \tag{83}$$

$$\widetilde{\delta U}^*_B \, \tilde{U}^*_B + \tilde{U}^*_B \, \widetilde{\delta U}^*_B = \widetilde{\delta X}^*_B. \tag{84}$$

Moreover, by (46) $(\tilde{U}, \tilde{X}, \tilde{S}) = (\tilde{U}^*, \tilde{X}^*, \tilde{S}^*)$ and the fact $\tilde{S}^*_{BN} = \tilde{U}^*_{BN} = 0$, we have

$$\tilde{U}^*_B \tilde{S}^*_B \tilde{U}^*_B = W_B,$$

which together with the assumption that $\|W - \nu I\| < \nu/\sqrt{2}$ for some $\nu > 0$ and the fact that $\tilde{U}^* \in \mathcal{U}^n_{++}$ implies

$$\|\tilde{U}^*_B \tilde{S}^*_B \tilde{U}^*_B - \nu I\| \;\; = \;\; \|W_B - \nu I\| < \nu/\sqrt{2}$$

and $(\tilde{U}^*_B,\ \tilde{S}^*_B) \in \mathcal{S}^{|B|}_{++} \times \mathcal{S}^{|B|}_{++}$. Using the conclusions above, relations (83) and (84), the first identity in (82), together with Lemma 2.4.5 and Lemma 2.4.6 with $H = 0$, we conclude that $\widetilde{\delta X}^*_B = \widetilde{\delta S}^*_B = \widetilde{\delta U}^*_B = 0$. Using similar arguments, we can also show that $\widetilde{\delta X}^*_N = \widetilde{\delta S}^*_N = \widetilde{\delta U}^*_N = 0$. ∎

As a consequence of the results obtained above, we have the following theorem when $W_{BN} = 0$.

Theorem 2.4.8 *If $W_{BN} = 0$ then the primal-dual weighted central path $(X(\nu), y(\nu), S(\nu))$ satisfies:*

i) $\lim_{\nu \downarrow 0} \sqrt{\nu}\ (\dot{X}(\nu), \dot{S}(\nu)) = 0$*;*

ii) $\|(X(\nu), S(\nu), y(\nu)) - (X^*, S^*, y^*)\| = \Theta(\nu)$*;*

ii) $\lim_{\nu \downarrow 0}(\dot{X}(\nu), \dot{S}(\nu), \dot{y}(\nu))$ *exists and (80) and (81) hold.*

Proof. Using Lemma 2.4.2 ii), Lemma 2.4.7 and the condition $W_{BN} = 0$, we obtain that $\tilde{X}^*_{BN} = \tilde{S}^*_{BN} = 0$, $\widetilde{\delta X}^*_B = 0$ and $\widetilde{\delta S}^*_N = 0$. Consequently, by Theorem 2.4.1, i) immediately follows. Statements ii) and iii) follow directly from i) and Theorem 2.4.4. ∎

2.5 Error bound analysis

By strengthening some of the results of the previous sections, in this section we derive a new error bound on the distance of a point lying in a certain neighborhood of the central path to the primal-dual optimal set.

For any given nonempty compact set $\mathcal{K} \subset \mathcal{G}_{++}$ and constants $\gamma, \tau > 0$, define

$$\mathcal{N}(\gamma, \tau, \mathcal{K}) \equiv \left\{(X, S, y) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m : \mathcal{G}(X, S, y) \in \tau\mathcal{K},\ \|X^{1/2} S X^{1/2} - \tau I\| \leq \gamma\tau\right\},$$

where the map $\mathcal{G}$ and the set $\mathcal{G}_{++}$ are defined in (26) and (27), respectively.

Observe that the set $\cup_{\tau>0}\mathcal{N}(\gamma,\tau,\mathcal{K})$ forms a neighborhood of the primal-dual central path. This neighborhood and related ones have been frequently used in the development of primal-dual interior point algorithms for SDP.

The following result gives a new error bound on the distance of a point lying in $\mathcal{N}(\gamma,\tau,\mathcal{K})$ to the primal-dual optimal set $\mathcal{F}_P^* \times \mathcal{F}_D^*$. Its proof will be given at the end of this section after we have derived stronger versions of the results of the previous sections.

Theorem 2.5.1 *Let $\gamma \in (0, 1/\sqrt{2})$ and any nonempty compact set $\mathcal{K} \subset \mathcal{G}_{++}$ be given. Then, there exists a constant $M = M(\gamma,\mathcal{K}) > 0$ such that*

$$\mathrm{dist}((X,S,y),\ \mathcal{F}_P^* \times \mathcal{F}_D^*) \leq M(\tau + \sqrt{\tau}\|W_{BN}\|), \tag{85}$$

for every $\tau \in (0,1]$ and $(X,S,y) \in \mathcal{N}(\gamma,\tau,\mathcal{K})$, where $W = W(X,S,\tau) \equiv X^{1/2}SX^{1/2}/\tau$.

In view of Proposition 2.2.1, for each $(\nu, W, \Delta C, \Delta b) \in (0,1] \times \mathcal{W} \times \mathcal{G}_{++}$, the system of nonlinear equations (22)-(24) has a unique solution, which in this section we denote by $(X(\nu,W,\Delta C,\Delta b), S(\nu,W,\Delta C,\Delta b), y(\nu,W,\Delta C,\Delta b))$ in order to emphasize and study its dependence on $(W,\Delta C,\Delta b)$. Moreover, in view of Theorem 2.3.2, the limit

$$\lim_{\nu\downarrow 0}(X(\nu,W,\Delta C,\Delta b), S(\nu,W,\Delta C,\Delta b), y(\nu,W,\Delta C,\Delta b)),$$

denoted by $(X(0,W,\Delta C,\Delta b), S(0,W,\Delta C,\Delta b), y(0,W,\Delta C,\Delta b))$, exists for any $(W,\Delta C,$ $\Delta b) \in \mathcal{W} \times \mathcal{G}_{++}$. Hence, the functions $X(\cdot,\cdot,\cdot,\cdot)$, $S(\cdot,\cdot,\cdot,\cdot)$ and $y(\cdot,\cdot,\cdot,\cdot)$ are well-defined in the set of points $[0,1] \times \mathcal{W} \times \mathcal{G}_{++}$. In an obvious way, we can also define the functions $\tilde{X}(t,W,\Delta C,\Delta b)$, $\tilde{S}(t,W,\Delta C,\Delta b)$ and $\tilde{y}(t,W,\Delta C,\Delta b)$ over the set of points $(t,W,\Delta C,\Delta b)$ $\in [0,1] \times \mathcal{W} \times \mathcal{G}_{++}$.

It turns out that re-parametrizations of the above functions are analytic according to the following definition. We say that that a function $f : \Omega \subseteq E \to F$, where E, F are two finite dimensional normed vector spaces, is analytic if there exists an open set $\mathcal{O} \subseteq E$ containing Ω and an analytic function $\tilde{f} : \mathcal{O} \to F$ such that $\tilde{f}$ restricted to Ω is equal to f.

Theorem 2.5.2 *There hold:*

i) the map $(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W} \times \mathcal{G}_{++} \to (\tilde{X}(t, W, \Delta C, \Delta b), \tilde{S}(t, W, \Delta C, \Delta b), \tilde{y}(t, W, \Delta C, \Delta b))$ *is analytic;*

ii) the map $(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W} \times \mathcal{G}_{++} \to (X(t^2, W, \Delta C, \Delta b), S(t^2, W, \Delta C, \Delta b), y(t^2, W, \Delta C, \Delta b))$ *is analytic.*

Proof. The proof of the theorem is identical to the proof of Theorem 2.3.11 and Theorem 2.3.2, except that when invoking the implicit function theorem, we should view $(t, W, \Delta C, \Delta b)$ as the parameter vector. ∎

Let

$$\mathcal{W}^b \equiv \{W \in \mathcal{W} \ : \ W_{BN} = 0\},$$

where $\mathcal{W}$ is defined in (28). One important result that we will establish next is that the function $(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}^b \times \mathcal{G}_{++} \to (X'(t^2, W, \Delta C, \Delta b), S'(t^2, W, \Delta C, \Delta b))$ is analytic. We emphasize that this result only holds over the smaller set $[0,1] \times \mathcal{W}^b \times \mathcal{G}_{++}$. Note also that this result does not follow immediately from Theorem 2.5.2(ii) since the derivative of the function in Theorem 2.5.2(ii) is not equal to the above function.

We now state a simple but crucial technical result needed to establish the result stated in the previous paragraph.

Proposition 2.5.3 *Let* $f : I \times E \to F$ *be a given analytic function, where* $I \subset \Re$ *is an interval and* E, F *are two finite dimensional normed vector spaces. Then, for any* $t^* \in I$*, the function* $g : I \times E \to F$ *defined as*

$$g(t,u) = \begin{cases} \frac{f(t,u) - f(t^*,u)}{t - t^*}, & \text{if } t \neq t^*; \\ \frac{\partial f}{\partial t}(t^*, u), & \text{if } t = t^*, \end{cases}$$

is analytic.

We are now ready to establish the result alluded to just before Proposition 2.5.3.

Lemma 2.5.4 *There hold:*

i) for any $(W, \Delta C, \Delta b) \in \mathcal{W}^b \times \mathcal{G}_{++}$, *we have*

$$\lim_{t \downarrow 0} \left(t \, \frac{\partial X}{\partial \nu}(t^2, W, \Delta C, \Delta b) \, , \, t \, \frac{\partial S}{\partial \nu}(t^2, W, \Delta C, \Delta b) \right) = 0;$$

ii) $(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}^b \times \mathcal{G}_{++} \to (X'(t^2, W, \Delta C, \Delta b), S'(t^2, W, \Delta C, \Delta b))$ *is analytic.*

Proof. In view of Theorem 2.4.8(i), we easily see that i) holds. Since partial derivatives of an analytic function are also analytic, it follows from Theorem 2.5.2(ii) that the functions

$$(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}^b \times \mathcal{G}_{++} \to \left(t \, \tfrac{\partial X}{\partial \nu}(t^2, W, \Delta C, \Delta b) \, , \, t \, \tfrac{\partial S}{\partial \nu}(t^2, W, \Delta C, \Delta b) \right),$$

$$(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}^b \times \mathcal{G}_{++} \to \left(\tfrac{\partial X}{\partial W}(t^2, W, \Delta C, \Delta b) \, , \, \tfrac{\partial S}{\partial W}(t^2, W, \Delta C, \Delta b) \right)$$

are both analytic. Using i) and Proposition 2.5.3, we conclude that the first function above divided by t is also analytic. We have thus shown that ii) holds. ∎

For $\gamma > 0$, let

$$\mathcal{W}(\gamma) \equiv \{ W \in \mathcal{S}^n_{++} : \|W - I\| \leq \gamma \}, \quad \mathcal{W}^b(\gamma) \equiv \{ W \in \mathcal{W}(\gamma) : W_{BN} = 0 \}.$$

We can easily see that if $\gamma < 1/\sqrt{2}$ then $\mathcal{W}(\gamma)$ and $\mathcal{W}^b(\gamma)$ are convex compact subsets of $\mathcal{W}$ and $\mathcal{W}^b$, respectively. For the remainder of this section, we let $\mathcal{K} \subset \mathcal{G}_{++}$ be any given nonempty compact set.

The next two results provide estimates on the sizes of the blocks of the matrices $X(\nu, W, \Delta C, \Delta b)$ and $S(\nu, W, \Delta C, \Delta b)$ first when $(W, \Delta C, \Delta b) \in \mathcal{W}^b(\gamma) \times \mathcal{K}$ and then for a general $(W, \Delta C, \Delta b) \in \mathcal{W}(\gamma) \times \mathcal{K}$.

Lemma 2.5.5 *Let* $\gamma \in (0, 1/\sqrt{2})$ *be given. Then, for all* $(\nu, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}^b(\gamma) \times \mathcal{K}$, *there holds*

$$\|(X(\nu, W, \Delta C, \Delta b), S(\nu, W, \Delta C, \Delta b)) - (X(0, W, \Delta C, \Delta b), S(0, W, \Delta C, \Delta b))\| = \mathcal{O}(\nu).$$

Proof. By the mean value theorem, we have

$$\|(X(\nu, W, \Delta C, \Delta b), S(\nu, W, \Delta C, \Delta b)) - (X(0, W, \Delta C, \Delta b), S(0, W, \Delta C, \Delta b))\|$$

$$\leq \sup_{\theta \in [0,1]} \|(X'(\theta\nu, W, \Delta C, \Delta b), S'(\theta\nu, W, \Delta C, \Delta b))\| \, \nu$$

By Theorem 2.5.4(ii) and the fact that $\mathcal{W}^b(\gamma)\times\mathcal{K}$ is compact, there exists a constant $M = M(\gamma,\mathcal{K}) > 0$ such that $\|(X'(\theta\nu,W,\Delta C,\Delta b),S'(\theta\nu,W,\Delta C,\Delta b))\| \le M$ for all $(\theta,\nu,W,\Delta C,$ $\Delta b)\in[0,1]\times[0,1]\times\mathcal{W}^b(\gamma)\times\mathcal{K}$. Hence, the lemma follows. ∎

Lemma 2.5.6 *For all* $(\nu,W,\Delta C,\Delta b)\in[0,1]\times\mathcal{W}(\gamma)\times\mathcal{K}$, *there hold*

$$X(\nu,W,\Delta C,\Delta b) - X(0,W^b,\Delta C,\Delta b) = \begin{pmatrix} \mathcal{O}(\|W_{BN}\|) & \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) \\ \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) & \mathcal{O}(\nu\|W_{BN}\|) \end{pmatrix} + \mathcal{O}(\nu) \quad (86)$$

$$S(\nu,W,\Delta C,\Delta b) - S(0,W^b,\Delta C,\Delta b) = \begin{pmatrix} \mathcal{O}(\nu\|W_{BN}\|) & \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) \\ \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) & \mathcal{O}(\|W_{BN}\|) \end{pmatrix} + \mathcal{O}(\nu) \quad (87)$$

where

$$W^b \equiv \begin{pmatrix} W_B & 0 \\ 0 & W_N \end{pmatrix}. \quad (88)$$

Proof. By Theorem 2.5.2(i), we know that $(\tilde{X}(t,W,\Delta C,\Delta b),\tilde{S}(t,W,\Delta C,\Delta b))$ is analytic over $[0,1]\times\mathcal{W}\times\mathcal{K}$. Hence, its derivative function $(\tilde{X}'(t,W,\Delta C,\Delta b),\tilde{S}'(t,W,\Delta C,\Delta b))$ is analytic, and hence continuous, over $[0,1]\times\mathcal{W}\times\mathcal{K}$. Since $[0,1]\times\mathcal{W}(\gamma)\times\mathcal{K}$ is compact, there exists a constant $M = M(\gamma,\mathcal{K}) > 0$ such that $\|(\tilde{X}'(\nu,W,\Delta C,\Delta b),\tilde{S}'(\nu,W,\Delta C,\Delta b))\| \le M$ for all $(t,W,\Delta C,\Delta b)\in[0,1]\times\mathcal{W}(\gamma)\times\mathcal{K}$. This together with (88) and the mean value theorem implies

$$\begin{aligned} &\|\tilde{X}(t,W,\Delta C,\Delta b) - \tilde{X}(t,W^b,\Delta C,\Delta b)\| \\ &= \sup_{\theta\in[0,1]} \|\tilde{X}'(t,\theta W+(1-\theta)W^b,\Delta C,\Delta b)\|\ \|W-W^b\| \\ &\le M\|W_{BN}\| = \mathcal{O}(\|W_{BN}\|), \end{aligned}$$

for all $(t,W,\Delta C,\Delta b)\in[0,1]\times\mathcal{W}(\gamma)\times\mathcal{K}$. This estimate and the fact that $\tilde{X}(t,W,\Delta C,\Delta b) = D_N(t)X(t^2,W,\Delta C,\Delta b)D_N(t)$ for all $t\in(0,1]$ and $(W,\Delta C,\Delta b)\in\mathcal{W}\times\mathcal{K}$ imply

$$\begin{aligned} &X(t^2,W,\Delta C,\Delta b) - X(t^2,W^b,\Delta C,\Delta b) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \left(\tilde{X}(t,W,\Delta C,\Delta b) - \tilde{X}(t,W^b,\Delta C,\Delta b)\right) \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{O}(\|W_{BN}\|) & \mathcal{O}(t\|W_{BN}\|) \\ \mathcal{O}(t\|W_{BN}\|) & \mathcal{O}(t^2\|W_{BN}\|) \end{pmatrix}. \end{aligned}$$

Noting that

$$\begin{aligned} X(t^2, W, \Delta C, \Delta b) - X(0, W^b, \Delta C, \Delta b) &= \Big(X(t^2, W, \Delta C, \Delta b) - X(t^2, W^b, \Delta C, \Delta b)\Big) \\ &\quad + \Big(X(t^2, W^b, \Delta C, \Delta b) - X(0, W^b, \Delta C, \Delta b)\Big) \end{aligned}$$

and using the above estimate together with Lemma 2.5.5, we immediately obtain (86) upon letting $\nu = t^2$. The estimate (87) can be proved in a similar way. ∎

We are now ready to state and prove the main result of this section.

Theorem 2.5.7 *For all $(\nu, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}(\gamma) \times \mathcal{K}$, there hold*

$$X(\nu, W, \Delta C, \Delta b) - X(0, W, \Delta C, \Delta b) = \begin{pmatrix} \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) & \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) \\ \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) & \mathcal{O}(\nu\|W_{BN}\|) \end{pmatrix} + \mathcal{O}(\nu), \quad (89)$$

$$S(\nu, W, \Delta C, \Delta b) - S(0, W, \Delta C, \Delta b) = \begin{pmatrix} \mathcal{O}(\nu\|W_{BN}\|) & \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) \\ \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) & \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\|) \end{pmatrix} + \mathcal{O}(\nu) \quad (90)$$

Proof. We will prove (89) only since the proof of (90) is similar. Since both $X(0, W, \Delta C, \Delta b)$ and $X(0, W^b, \Delta C, \Delta b)$ are in $\mathcal{F}_P^*$, we have

$$X_{BN}(0, W, \Delta C, \Delta b) = X_{BN}(0, W^b, \Delta C, \Delta b) = 0,$$

and $X_N(0, W, \Delta C, \Delta b) = X_N(0, W^b, \Delta C, \Delta b) = 0$ for any $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{K}$. Hence, in view of Lemma 2.5.6, it suffices to show that for all $(\nu, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}(\gamma) \times \mathcal{K}$, we have

$$X_B(\nu, W, \Delta C, \Delta b) - X_B(0, W, \Delta C, \Delta b) = \mathcal{O}(\sqrt{\nu}\,\|W_{BN}\| + \nu). \quad (91)$$

Indeed, using the fact that $\tilde{X}_B(t, W, \Delta C, \Delta b)$ is analytic over the compact set $[0,1] \times \mathcal{W}(\gamma) \times \mathcal{K}$ due to Theorem 2.5.2(i), we conclude that

$$\begin{aligned} X_B(t^2, W, \Delta C, \Delta b) - X_B(0, W, \Delta C, \Delta b) &= \tilde{X}_B(t, W, \Delta C, \Delta b) - \tilde{X}_B(0, W, \Delta C, \Delta b) \\ &= \frac{\partial}{\partial t}\tilde{X}_B(0, W, \Delta C, \Delta b)t + \mathcal{O}(t^2), \end{aligned} \quad (92)$$

for every $(t, W, \Delta C, \Delta b) \in [0,1] \times \mathcal{W}(\gamma) \times \mathcal{K}$. Moreover, by Lemma 2.4.7, we have

$$\frac{\partial \tilde{X}_B}{\partial t}(0, W^b, \Delta C, \Delta b) = 0,$$

for any $(W, \Delta C, \Delta b) \in \mathcal{W} \times \mathcal{K}$, where W^b is defined in (88). Hence, for every $(W, \Delta C, \Delta b) \in \mathcal{W}(\gamma) \times \mathcal{K}$, we have

$$\frac{\partial}{\partial t}\tilde{X}_B(0, W, \Delta C, \Delta b) = \frac{\partial}{\partial t}\tilde{X}_B(0, W^b, \Delta C, \Delta b) + \mathcal{O}(\|W - W^b\|) = \mathcal{O}(\|W_{BN}\|). \tag{93}$$

Combining (92) and (93), we obtain (91) upon letting $\nu = t^2$. ∎

The proof of Theorem 2.5.1 now follows from Assumption **A.1** and Theorem 2.5.7 with $\nu = \tau$, $W = X^{1/2}SX^{1/2}/\tau$, $(X, S) = (X(\nu, W, \Delta C, \Delta b), S(\nu, W, \Delta C, \Delta b))$ and the fact $(X(0, W, \Delta C, \Delta b), S(0, W,$
$\Delta C, \Delta b), y(0, W, \Delta C, \Delta b)) \in \mathcal{F}_P^* \times \mathcal{F}_D^*$.

2.6 Superlinear convergence criteria

In this section, we consider a sufficient condition introduced by Potra and Sheng [103], which guarantees the superlinear convergence of a class of primal-dual interior point algorithms for SDP, and show that it is equivalent to a natural condition about the matrix $W(X, S, \tau)$ of Theorem 2.5.1.

For sake of concreteness, we will focus our attention on the algorithm and results obtained in Potra and Sheng (see Algorithm 3.1 in [104]), but we remark that our discussion also applies to a broader class of algorithms. Potra and Sheng [104] have developed a primal-dual infeasible-interior-point algorithm which, for some $\alpha \in (0, 1/2]$, generates a sequence of iterates $\{(X^k, S^k, y^k)\} \subseteq \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \Re^m$ satisfying

$$\|W^k - I\|_F \leq \alpha, \quad r_p^k = \frac{\tau_k}{\tau_0} r_p^0, \quad r_d^k = \frac{\tau_k}{\tau_0} r_d^0, \tag{94}$$

for some sequence $\{\tau_k\} \subset \Re_{++}$ converging to 0 at least Q-linearly, where

$$\begin{aligned} r_p^k &\equiv \mathcal{A}X^k - b, \\ r_d^k &\equiv \mathcal{A}^* y^k + S^k - C, \\ W^k &\equiv \frac{(X^k)^{1/2} S^k (X^k)^{1/2}}{\tau_k}, \end{aligned}$$

for all $k \geq 0$. The derived linear rate of convergence of the sequence $\{\tau_k\}$ is sufficient to guarantee polynomial convergence of their method under some suitable conditions on

the initial point (X^0, S^0, y^0). Observe that the first condition in (94) implies that $\tau_k = \Theta(X^k \bullet S^k/n)$, and hence asymptotic convergence of $\{X^k \bullet S^k/n\}$ can be derived from the one obtained for $\{\tau_k\}$.

Some sufficient conditions have been developed in the literature which guarantee the Q-superlinear convergence of $\{\tau_k\}$ to zero. One such condition is the tangential condition proposed by Kojima et al. [62], namely

$$\lim_{k\to\infty} W^k = I. \tag{95}$$

Another such condition, and the one which will be the main subject of this section, is the one that has been proposed by Potra and Sheng [103], namely

$$\lim_{k\to\infty} X^k S^k/\sqrt{\tau_k} = 0. \tag{96}$$

We remark that Potra and Sheng [103]) have shown that the tangential condition (95) implies their condition (96). Moreover, they have also established the following superlinear convergence result.

Proposition 2.6.1 (Theorem 6.1 of [103]) *If (96) holds, then the sequence $\{\tau_k\}$ generated by Algorithm 3.1 of [104] converges to zero Q-superlinearly. Moreover, if $X^k S^k = \mathcal{O}(\tau_k^{0.5+\sigma})$ for some constant $\sigma > 0$, then $\{\tau_k\}$ converges to zero with Q-order at least $1 + \min\{\sigma, 0.5\}$.*

A natural relaxation of the tangential condition (95) is the condition that

$$\lim_{k\to\infty} W^k_{BN} = 0. \tag{97}$$

Surprisingly, the following result shows that it is equivalent to Potra and Sheng's condition (96).

Proposition 2.6.2 *Let $\theta_k \equiv \|X^k S^k\|/\sqrt{\tau_k}$. Then, $\|W^k_{BN}\| + \sqrt{\tau_k} = \Theta(\theta_k + \sqrt{\tau_k})$.*

Proof. We first show that $\|W^k_{BN}\| + \sqrt{\tau_k} = \mathcal{O}(\theta_k + \sqrt{\tau_k})$. By Lemma 4.2 of [103], we have

$$X^k = \begin{pmatrix} \Theta(1) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\tau_k) \end{pmatrix}, \quad S^k = \begin{pmatrix} \mathcal{O}(\tau_k) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \Theta(1) \end{pmatrix}.$$

and hence

$$\begin{aligned}\frac{X^kS^k}{\sqrt{\tau_k}} &= \frac{1}{\sqrt{\tau_k}}\begin{pmatrix} \Theta(1) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\tau_k) \end{pmatrix}\begin{pmatrix} \mathcal{O}(\tau_k) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \Theta(1) \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(1) \\ \mathcal{O}(\tau_k) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}.\end{aligned}$$

According to the definition of θ_k , we then conclude that

$$\frac{X^kS^k}{\sqrt{\tau_k}} = \begin{pmatrix} \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\theta_k) \\ \mathcal{O}(\tau_k) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}. \tag{98}$$

By Lemma 4.5 of [103], we have

$$(X^k)^{1/2} = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}, \quad (X^k)^{-1/2} = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/\sqrt{\tau_k}) \end{pmatrix}, \tag{99}$$

which together with (98) imply

$$\begin{aligned}W^k &= \frac{(X^k)^{1/2}S^k(X^k)^{1/2}}{\tau_k} = \frac{1}{\sqrt{\tau_k}}(X^k)^{-1/2}\left(\frac{X^kS^k}{\sqrt{\tau_k}}\right)(X^k)^{1/2} \\ &= \frac{1}{\sqrt{\tau_k}}\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/\sqrt{\tau_k}) \end{pmatrix}\begin{pmatrix} \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\theta_k) \\ \mathcal{O}(\tau_k) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}.\end{aligned}$$

Since the (B,N)-block of the matrix in the right hand side of the above identity is $\mathcal{O}(\theta_k + \sqrt{\tau_k})$, we conclude that $\|W^k_{BN}\| + \sqrt{\tau_k} = \mathcal{O}(\theta_k + \sqrt{\tau_k})$.

Next we show that $\theta_k + \sqrt{\tau_k} = \mathcal{O}(\|W^k_{BN}\| + \sqrt{\tau_k})$. By the first condition in (94), we have

$$W^k = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\|W^k_{BN}\|) \\ \mathcal{O}(\|W^k_{BN}\|) & \mathcal{O}(1) \end{pmatrix},$$

which together with (99) implies that

$$\begin{aligned}\frac{X^kS^k}{\sqrt{\tau_k}} &= \sqrt{\tau_k}\,(X^k)^{1/2}\left(\frac{(X^k)^{1/2}S^k(X^k)^{1/2}}{\tau_k}\right)(X^k)^{-1/2} = \sqrt{\tau_k}\,(X^k)^{1/2}W^k(X^k)^{-1/2} \\ &= \sqrt{\tau_k}\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\tau_k}) \\ \mathcal{O}(\sqrt{\tau_k}) & \mathcal{O}(\sqrt{\tau_k}) \end{pmatrix}\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\|W^k_{BN}\|) \\ \mathcal{O}(\|W^k_{BN}\|) & \mathcal{O}(1) \end{pmatrix}\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/\sqrt{\tau_k}) \end{pmatrix}\end{aligned}$$

$$= \begin{pmatrix} \mathcal{O}\left(\sqrt{\tau_k} + \sqrt{\tau_k}\,\|W_{BN}^k\|\right) & \mathcal{O}\left(\sqrt{\tau_k} + \|W_{BN}^k\|\right) \\ \mathcal{O}\left(\tau_k + \tau_k\|W_{BN}^k\|\right) & \mathcal{O}\left(\sqrt{\tau_k} + \sqrt{\tau_k}\,\|W_{BN}^k\|\right) \end{pmatrix}.$$

This together with the definition of θ_k implies that $\theta_k + \sqrt{\tau_k} = \mathcal{O}(W_{BN}^k + \sqrt{\tau_k})$. ∎

In view of the equivalence between (96) and (97), it follows that a sufficient condition for the superlinear convergence of the path-following algorithm outlined in this section is that the sequence of matrices $\{W^k\}$ approaches the set of block diagonal matrices. Clearly, this a much weaker condition than (95), which requires this sequence to approach the identity matrix.

2.7 Concluding remarks

In this section we provide some final remarks related to the results derived in this chapter.

Under the assumptions of this chapter, we have shown that the re-parametrized $(W, \Delta C, \Delta b)$-weighted central path $(X(t^2), S(t^2), y(t^2))$ is analytic at $t = 0$ and that the condition $W_{BN} = 0$ implies that $\lim_{\nu \downarrow 0}(\dot{X}(\nu), \dot{S}(\nu), \dot{y}(\nu))$ exists. Based on the latter conclusion, it is natural to wonder whether the path $(X(\nu), S(\nu), y(\nu))$ is analytic at $\nu = 0$ when W is block-diagonal. Note that the answer to this question is affirmative when $(W, \Delta C, \Delta b) = (I, 0, 0)$, i.e., the weighted central path is exactly the central path (see Halická [50]).

In this chapter, we have proved that the rate of convergence of the $(W, \Delta C, \Delta b)$-weighted central path $(X(\nu), S(\nu), y(\nu))$ towards the optimal solution set is $\mathcal{O}(\sqrt{\nu})$ (and $\mathcal{O}(\nu)$ when $W_{BN} = 0$). In contrast, Preiß and Stoer [107] have shown that the rate of convergence of the weighted central paths associated with the map $(XS+SX)/2$ is always $\mathcal{O}(\nu)$(see also Lu and Monteiro [70]). An error bound of this type has also been shown by Kojima et al. [63], where it is shown that an interior-point algorithm based on a centering condition associated with the $(XS+SX)/2$-map does not need to approach the central path tangentially in order to converge superlinearly. On the other hand, the iterates of all superlinearly-convergent interior-point algorithms based on centering conditions associated with the map $X^{1/2}SX^{1/2}$ that have been proposed in the literature are required to approach the central path tangentially. The latter requirement is natural in view of the fact that it forces $(X^k)^{1/2}S^k(X^k)^{1/2}$

to approach a block diagonal matrix (the identity matrix), and hence it reduces the bound on the distance of (X^k, S^k, y^k) to the optimal solution set from the usual $\mathcal{O}(\sqrt{\mu_k})$ to $o(\sqrt{\mu_k})$ (see Theorem 2.5.1).

CHAPTER III

LARGE-SCALE SEMIDEFINITE PROGRAMMING VIA A SADDLE POINT MIRROR-PROX ALGORITHM

3.1 Preliminary Remarks

Consider a semidefinite program

$$\min_x \{\mathrm{Tr}(cx) : x \in \mathcal{N} \cap \mathbf{S}_+\}, \tag{100}$$

where $\mathbf{S}_+$ is the cone of positive semidefinite matrices in the space $\mathbf{S}$ of symmetric block-diagonal matrices with a given block-diagonal structure, $\mathcal{N}$ is an affine subspace in $\mathbf{S}$ and $c \in \mathbf{S}$. The goal of this chapter is to investigate the possibility of utilizing favourable sparsity patterns of a large-scale problem (100) (that is, the sparsity pattern of diagonal blocks in matrices from $\mathcal{N}$) when solving the problem by a simple first-order method. To motivate our goal, let us start with discussing whether it makes sense to solve (100) by first-order methods, given the breakthrough developments in the theory and implementation of Interior Point methods (IPMs) for Semidefinite Programming (SDP) we have witnessed during the last decade. Indeed, IPMs are polynomial time methods and as such allow to solve SDPs within accuracy ϵ at a low iteration count (proportional to $\ln(1/\epsilon)$) and thus capable of producing high-accuracy solutions. Note, however, that IPMs are Newton-type methods, with an iteration which requires assembling and solving a Newton system of n linear equations with n unknowns, where $n = \min[\dim\mathcal{N}, \mathrm{codim}\,\mathcal{N}]$ is the minimum of the design dimensions of the problem and its dual. Typically, the Newton system is dense, so that the cost of solving it by standard Linear Algebra techniques is $O(n^3)$ arithmetic operations. It follows that in reality the scope of IPMs in SDP is restricted to problems with n at most few thousands – otherwise a single iteration will "last forever". At the present level of our knowledge, the only way to process numerically SDPs with n of order of 10^4 or more seems to use simple first-order optimization techniques with computationally cheap iterations. Although

all known first-order methods in the large-scale case exhibit slow – sublinear – convergence and thus are unable to produce high-accuracy solutions in realistic time, medium-accuracy solutions are still achievable. Historically, the first SDP algorithm of the latter type was the *spectral bundle* method [54] – a version of the well-known bundle method for nonsmooth convex minimization "tailored" to semidefinite problems. A strong point of the present method is in its modest requirements on our abilities to handle matrices from $\mathcal{N}$ – all we need is to compute few largest eigenvalues and associated eigenvectors of such matrices. This task can be carried out routinely when the largest size ζ of diagonal blocks in matrices from $\mathbf{S}$ is not too large, say, $\zeta \leq 1000$. Note that under this limitation, n still can be of order of 10^5, meaning that (100) is far beyond the scope of IPMs. Moreover, the task in question still can be carried out when ζ is much larger than the above limit, provided that diagonal blocks in the matrices $A \in \mathcal{N}$ possess favourable sparsity patterns. A weak point of the spectral bundle method, at least from the theoretical viewpoint, is the convergence rate: the inaccuracy in terms of the objective can decrease with the iteration count t as slowly as $O(t^{-1/2})$ (this is the best possible, in the large scale case, rate of convergence of first-order methods on nonsmooth convex programs). Also, theoretical convergence rate results are not established for the first-order SDP algorithms proposed recently in [20, 19]. Recently, novel $O(t^{-1})$-converging first-order algorithms, based on smooth saddle-point reformulation of nonsmooth convex programs were developed [92, 91, 90]. Numerical results presented in these papers (including those on genuine SDP with n as large as 100,000 – 190,000 [90]) demonstrate high computational potential of the proposed methods. However, theoretical and computational advantages exhibited by the $O(t^{-1})$-converging methods as compared to algorithms like spectral bundle have their price, specifically, the necessity to operate with eigenvalue decompositions of the matrices from $\mathbf{S}$ rather than computing a few largest eigenvalues of matrices from $\mathcal{N}$. As a result, the algorithms from [92, 91, 90] as applied to (100) become impractical, when the largest size ζ of diagonal blocks in the matrices from $\mathbf{S}$ exceeds about 1000.

The goal of this chapter is to demonstrate that one can extend the scope of $O(t^{-1})$-converging first-order methods as applied to semidefinite program (100) beyond the just

outlined limits by assuming that diagonal blocks in the matrices from $\mathcal{N}$ possess favourable sparsity patterns. This type of semidefinite program (100) has also been studied in [38] via matrix completion in the context of IPM. The outline of this chapter is as follows. In Section 3.2, we explain what a "favourable sparsity pattern" is and introduce some notation and definitions which will be used throughout this chapter. In Section 3.3, we develop our main tool, specifically, demonstrate that positive semidefiniteness of a large symmetric matrix A possessing a favourable sparsity pattern can be represented via positive semidefiniteness of a bunch of smaller matrices linked, in a linear fashion, to A. We derive also the "dual counterpart" of the outlined representation, which expresses the possibility of positive semidefinite completion of a "well-structured" partially defined symmetric matrix in terms of positive semidefiniteness of a specific bunch of fully defined submatrices of the matrix[1)]. In Section 3.4 we utilize the aforementioned representations to derive saddle point formulations of some large-scale SDP problems, specifically, those of computing Lovász capacity of a graph and the MAXCUT problem, with emphasis on the case when the incidence matrix of the underlying graph possesses a favourable sparsity pattern. We demonstrate that the complexity of solving these problems within a fixed relative accuracy by an appropriate $O(t^{-1})$-converging first-order method (namely, the Mirror-Prox algorithm from [90]) is by orders of magnitude less than complexity associated with IPMs, and show that with our approach, we indeed can utilize a favourable sparsity pattern in the incidence matrix. In concluding Section 3.5, we illustrate our constructions by numerical results for the MAXCUT and Lovász capacity problems on well-structured sparse graphs.

3.2 Well-structured sparse symmetric matrices

In this section, we motivate and define the notion of a symmetric matrix with "favourable sparsity pattern" and introduce notation to be used throughout this chapter.

Motivation. To get an idea what a "favourable sparsity pattern" might be, consider the semidefinite program (100), and let A^ℓ, $\ell = 1, ..., L$, be the diagonal blocks of a generic

[1)] This result, which we get "for free", can be also obtained, with a moderate effort, from general results of [46] on existence of positive semidefinite completions.

matrix from $\mathcal{N}$. Assume that these blocks possess certain sparsity patterns. How could we utilize this sparsity? Our first observation is that even high sparsity by itself can be of no use. Indeed, consider the simplest SDP-related computational issue, that is, checking whether a matrix $A = \mathrm{Diag}\{A^1, ..., A^L\}$ from $\mathcal{N}$ is positive semidefinite. Assuming that we are checking positive semidefiniteness of sparse symmetric matrices A^ℓ by applying Cholesky factorization algorithm, that is, by trying to represent A^ℓ as $D_\ell D_\ell^T$ with lower triangular D_ℓ, the nonzeros in D_ℓ will, generically, be the entries i, j with $i - v_i \leq j \leq i$, where $i - v_i = \min\{j : A^\ell_{ij} \neq 0\}$. In other words, when adding to the original pattern of nonzero entries all entries i, j with $i - v_i \leq j \leq i$ (and all symmetric entries), we do not alter the fill in of the Cholesky factor. Therefore, we do not lose much by assuming that the original pattern of nonzeros already was comprised of all sub-diagonal entries (i, j) with $i - v_i \leq j \leq i$, with added symmetric entries. For notational convenience, we prefer to work with "symmetric" situation, where the nonzero entries are super-diagonal entries i, j with $i \leq j \leq i + v_i$, and the symmetric sub-diagonal entries; note that matrices of the former type can be obtained from those of the latter one by reversing the order of rows and columns. We arrive at the notion of a *well-structured* sparse $n \times n$ symmetric matrix with sparsity pattern given by a nonnegative integral vector v_i such that $i + v_i \leq n$ for all i; the "hard zero" super-diagonal entries i, j $(i \leq j)$ in such a matrix are those with $j > i + v_i$. Note that for such a matrix A, the "hard zeros" in the *upper triangular* factor U of the Cholesky factorization $A = UU^T$, are exactly the same as hard zeros in the upper triangular part of A. In particular, if A is a well-structured sparse symmetric matrix with $\sum_i v_i << n^2$, then it is relatively easy to check whether or not $A \succeq 0$; to this end, it suffices to apply to A the Cholesky factorization algorithm (where the factorization being sought is $A = UU^T$ with upper triangular U).

Next, we introduce terminology and notation for dealing with "well-structured", in the sense we have just motivated, sparsity patterns.

Simple sparsity structures and associated entities. Let $v \in \mathbf{R}^n$ be a *simple sparsity structure* – a nonnegative integral vector such that $i + v_i \leq n$ for all $i \leq n$. We associate

with structure v the following entities:

1. A subspace $\mathbf{S}^{(v)}$ in the space $\mathbf{S}^n$ of symmetric $n \times n$ matrices; $\mathbf{S}^{(v)}$ is comprised of all matrices $[A_{ij}]_{i,j=1}^n$ from $\mathbf{S}^n$ such that $A_{ij} = 0$ for $j > i + v_i$.

2. The set $I = \{i_1 < i_2 < ... < i_m\}$ of all integers representable as $i + v_i$ with $i \leq n$. Note that $i_m = n$, since $n + v_n = n$ (recall that $i + v_i \leq n$ and $v_i \geq 0$). We refer to m as the *number of blocks* in v.

3. The sets

$$J_k = \{i \leq i_k : i + v_i \geq i_k\},\ J'_k = \{i \in J_k : i \leq i_{k-1}\},\ k = 1, ..., m, \tag{101}$$

   where $i_0 = 0$ (that is, $J'_1 = \emptyset$). Note that $J_k \backslash J_{k-1} = \{i_{k-1} + 1, ..., i_k\}$ and that $J'_k = J_{k-1} \cap J_k$, where $J_0 = \emptyset$.

4. The set of *occupied* cells ij – those with $i \leq j \leq i + v_i$. For an occupied cell ij, both integers $i + v_i$ and $j + v_j$ are elements of the set $I = \{i_1, ..., i_m\}$; thus, $\min[i + v_i, j + v_j] = i_{k_+}$ for certain $k_+ = k_+(i, j) \leq m$. Since $j \leq i + v_i$, we have $j \leq \min[i + v_i, j + v_j] = i_{k_+}$. Therefore the smallest k, let it be called $k_- = k_-(i, j)$, such that $j \leq i_k$, satisfies $k_- \leq k_+$. Since $j + v_j$ is one of i_s, we conclude that $j + v_j \geq i_{k_-}$. Note that the segment $D_{ij} = \{k_-, k_- + 1, ..., k_+\}$ is exactly the segment of those k for which i and j belong to J_k; we denote by $\ell(i, j)$ the cardinality of D_{ij}.

5. Two diagonal matrices $\mathcal{L}$ and $\mathcal{K}$ defined as

$$\mathcal{L} = \mathrm{Diag}\{\ell(1, 1)^{-1/2}, ..., \ell(n, n)^{-1/2}\},\ \mathcal{K} = \mathrm{Diag}\{\ell(1, 1), ..., \ell(n, n)\}. \tag{102}$$

We now provide an example to illustrate the definitions given above. Consider a subspace of $\mathbf{S}^7$, consisting of all symmetric matrices with nonzero entries specified as below

$$\begin{bmatrix} * & * & * & * & & & \\ * & * & * & & & & \\ * & * & * & * & * & * & \\ * & & * & * & * & & \\ & & * & * & * & * & * \\ & & * & & * & * & * \\ & & & & * & * & * \end{bmatrix} \in \mathbf{S}^7.$$

We observe that the subspace defined above is $\mathbf{S}^{(v)}$ with $v = (3, 1, 3, 1, 2, 1, 0)^T$. We easily see that $m = 5$, and $i_1 = 3$, $i_2 = 4$, $i_3 = 5$, $i_4 = 6$ and $i_5 = 7$, and hence $I = \{3, 4, 5, 6, 7\}$. Using (101), we have

$$J_1 = \{1, 2, 3\}, \quad J_2 = \{1, 3, 4\}, \quad J_3 = \{3, 4, 5\}, \quad J_4 = \{3, 5, 6\}, \quad J_5 = \{5, 6, 7\},$$

$$J_1' = \emptyset, \quad J_2' = \{1, 3\}, \quad J_3' = \{3, 4\}, \quad J_4' = \{3, 5\}, \quad J_5' = \{5, 6\}.$$

Using the definition of D_{ij}, we have

$$D_{11} = \{1, 2\}, \quad D_{22} = \{1\}, \quad D_{33} = \{1, 2, 3, 4\}, \quad D_{44} = \{2, 3\},$$

$$D_{55} = \{3, 4, 5\}, \quad D_{66} = \{4, 5\}, \quad D_{77} = \{5\},$$

and hence,

$$\ell(1,1) = 2, \quad \ell(2,2) = 1, \quad \ell(3,3) = 4, \quad \ell(4,4) = 2, \quad \ell(5,5) = 3, \quad \ell(6,6) = 2, \quad \ell(7,7) = 1.$$

Therefore, we obtain that

$$\mathcal{L} = \mathrm{Diag}\left\{\frac{1}{\sqrt{2}},\ 1,\ \frac{1}{2},\ \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{2}},\ 1\right\}, \quad \mathcal{K} = \mathrm{Diag}\{2,\ 1,\ 4,\ 2,\ 3,\ 2,\ 1\}.$$

Finally, in the sequel $\lambda_{\min}(A)$ (resp., $\lambda_{\max}(A)$) denotes the minimal (resp., maximal) eigenvalue of a symmetric matrix A. δ_j^i denotes the Kronecker delta. For a finite set $\mathcal{J}$, we denote its cardinality by $|\mathcal{J}|$.

3.3 Representation results for well-structured sparse symmetric matrices

Consider again the semidefinite program (100). Assuming that the diagonal blocks A^ℓ in a generic matrix $A \in \mathcal{N}$ to be sparse, with well-structured sparsity pattern as defined in Section 3.2, it is relatively easy to verify whether the Linear Matrix Inequalities (LMIs) are satisfied at a given point (since the Cholesky factorization $A^\ell = U_\ell U_\ell^T$ with upper triangular U_ℓ does not increase fill in). This possibility, however, in many respects is not sufficient. When solving SDPs by numerous advanced methods, including interior point ones, we would prefer to deal with many small dense LMIs rather than with few large sparse ones, at least in the case when the total row size of the former system of LMIs is of the same order of magnitude as the total row size of the latter system. In this respect, the following question is of definite interest:

> Given a well-structured sparse matrix A, is it possible to express the fact that $A \succeq 0$ by a system of relatively small LMIs in variables A_{ij} and perhaps additional variables?

We are about to give an affirmative answer to this question.

3.3.1 Positive semidefiniteness of well-structured sparse matrices

In this subsection, we will provide some necessary and sufficient conditions for a matrix from $\mathbf{S}^{(v)}$ to be positive semidefinite. The following notations will be used throughout the remaining chapter.

Notation. Let $J \subset \{1, ..., n\}$ be an index set with $\ell > 0$ elements. We denote by $[B_{ij}]_{i,j\in J}$ the $\ell \times \ell$ matrix obtained from B by extracting the rows and columns with indices in J, and by $]B_{ij}[_{i,j\in J}$ – the $n \times n$ matrix with entries B_{ij} for all $i, j \in J$ and zero entries for the remaining pairs i, j.

The following notations will be used in this subsection and Subsection 3.4.2 of this chapter.

Let $v \in \mathbf{R}^n$ be a simple sparsity structure, and J_k, $k = 1, ..., m$, be the corresponding index sets (see Section 3.2). We define $\mathbf{B}$ as an Euclidean space comprised of collections $B = \{B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k}\}_{k=1}^m$ of symmetric matrices, i.e.,

$$\mathbf{B} \equiv \left\{B = (B_1, \ldots, B_m) : B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k},\ k = 1, \ldots, m\right\},$$

and equipped with natural linear operations and the norm

$$\|B\|_F = \sqrt{\sum_{k=1}^m \|B_k\|_F^2},$$

where $\|B_k\|_F$ is the Frobenius norm of B_k. For $B = \{B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k}\}_{k=1}^m \in \mathbf{B}$, we set

$$B^k =]B^k_{ij}[_{i,j\in J_k} \in \mathbf{S}^{(v)},\ k = 1, ..., m.$$

and define the linear mapping $\mathcal{S}(B) : \mathbf{B} \to \mathbf{S}^{(v)}$ as

$$\mathcal{S}(B) = \sum_{k=1}^m B^k.$$

Proposition 3.3.1 (i) *A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if there exists $B = \{B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k} \succeq 0\}_{k=1}^m \in \mathbf{B}$ such that*

$$A = \mathcal{S}(B) \equiv \sum_{k=1}^m B^k. \tag{103}$$

(ii) *Whenever $B = \{B_k = [B^k_{ij}]_{i,j\in J_k} \succeq 0\}_{k=1}^m$ satisfies* (103)*, one has, for any $n \times n$ real matrix W,*

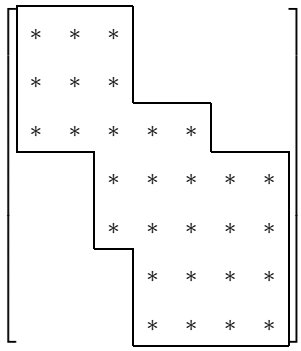
$$\sum_{k=1}^m \|W^T B^k W\|_F^2 \le \|W^T A W\|_F^2. \tag{104}$$

(iii) *We have*

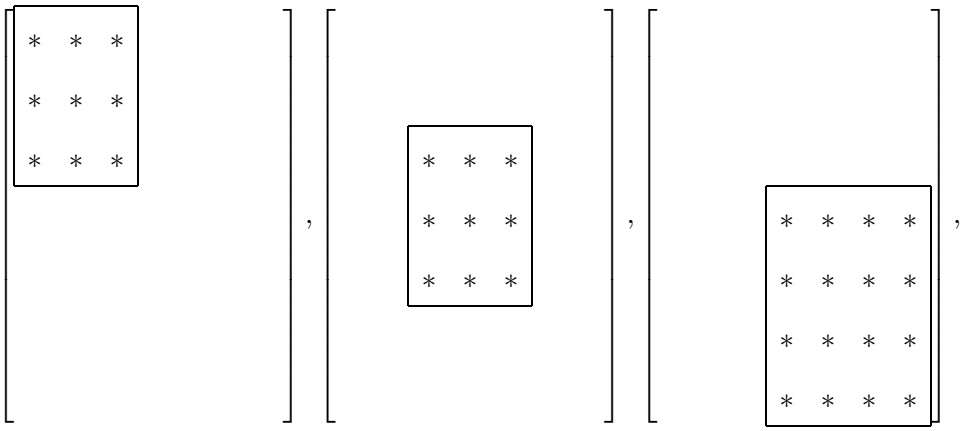
$$\forall B \in \mathbf{B} : \|\mathcal{L}^{1/2}\mathcal{S}(B)\mathcal{L}^{1/2}\|_F \le \|B\|_F, \tag{105}$$

where $\mathcal{L}$ is given by (102)*.*

Illustration: Overlapping block-diagonal structure. Before proving Proposition 3.3.1, it makes sense to "visualize" its simplest "overlapping block-diagonal" version. Consider a symmetric block-matrix of the form

$A =$ 

where $*$ mark nonzero blocks. Proposition 3.3.1.(i) says that such a matrix is positive semidefinite if and only if it is the sum of positive semidefinite matrices of the form



and similarly when the number of overlapping diagonal blocks is > 3.

Proof of Proposition 3.3.1. (i): Induction in m. For $m = 1$ the statement is evident. Assuming that the statement is valid for $m = s$, let us prove it for $m = s + 1$. The "if" part is evident; thus, assume that $A \in \mathbf{S}^{\nu}$ is $\succeq 0$, and let us prove the existence of the required B_k. For $\epsilon \geq 0$, let $A_\epsilon \equiv A + \epsilon I = \left[\begin{array}{c|c} P & Q \\ \hline Q^T & R + \epsilon I \end{array}\right]$ with $i_{m-1} \times i_{m-1}$ block P. For $\epsilon > 0$, let $B^\epsilon = \left[\begin{array}{c|c} Q(R+\epsilon I)^{-1}Q^T & Q \\ \hline Q^T & R + \epsilon I \end{array}\right]$, so that $B^\epsilon \succeq 0$. By the Schur Complement Lemma, we have $A_\epsilon - B^\epsilon \succeq 0$, thus, B^ϵ remains bounded as $\epsilon \to +0$. Thus, we can find a

sequence $\epsilon_t \to +0$, $t \to \infty$, and a matrix B^m such that

$$B^m = \lim_{t\to\infty} B^{\epsilon_t}. \tag{106}$$

Observe that both B^m and $A - B^m$ are $\succeq 0$. By construction, $B^m =]B^m_{ij}[_{i,j\in J_m}$; besides this, the rows i and the columns j in $C = A - B^m$ with $i, j > i_{m-1}$ are zero. Removing these rows and columns, we get an $i_{m-1} \times i_{m-1}$ matrix $\bar{C} \in \mathbf{S}^{(v')}$, where $v'_i = \min[i_{m-1} - i, v_i]$, $1 \le i \le i_{m-1} = \dim v'$. Clearly, the number of blocks in v' is $m-1$, and the corresponding index sets J_k, $1 \le k \le m-1$, are the same as for v. Applying to $\bar{C}$ the inductive hypothesis, we can find $m-1$ matrices $B^k =]B^k_{ij}[_{i,j\in J_k} \succeq 0$, $k = 1, ..., m-1$, such that $C = \sum_{k=1}^{m-1} B^k$, whence $A = C + B^m = \sum_{k=1}^{m} B^k$ with $B^k \succeq 0$ of the required structure. The induction is over.

(ii): For matrices $B, C \succeq 0$, one has $\mathrm{Tr}(BC) \ge 0$. It follows that under the premise of (ii) one has $\|\sum_k W^T B^k W\|_F^2 \ge \sum_k \|W^T B^k W\|_F^2$.

(iii): Let $A = \mathcal{S}(B)$, so that $A_{ij} = \sum_{k:i,j\in J_k} B^k_{ij}$. Recall from Section 3.2 that $\ell(i,j) = |\{k : i, j \in J_k\}|$ for every i, j. We have

$$\begin{aligned}
\|\mathcal{L}^{1/2} A \mathcal{L}^{1/2}\|_F^2 &= \sum_{i,j} A_{ij}^2 \ell^{-1/2}(i,i)\ell^{-1/2}(j,j) \\
&= \sum_{i,j} \left(\sum_{k:i,j\in J_k} B^k_{ij} \right)^2 \ell^{-1/2}(i,i)\ell^{-1/2}(j,j) \\
&\le \sum_{i,j} \sum_{k:i,j\in J_k} (B^k_{ij})^2 \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \\
&\le \left(\max_{i,j} \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \right) \sum_{i,j} \sum_{k:i,j\in J_k} (B^k_{ij})^2 \\
&= \left(\max_{i,j} \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \right) \|B\|_F^2;
\end{aligned}$$

thus, in order to prove (iii) it suffices to verify that

$$\ell(i,j) \le \sqrt{\ell(i,i)\ell(j,j)}$$

for every i, j. This is evident due to

$$\ell(i,j) = |\{k : i, j \in J_k\}| \le \min\left[|\{k : i \in J_k\}|, |\{k : j \in J_k\}|\right] = \min[\ell(i,i), \ell(j,j)].$$

■

Proposition 3.3.1(i) establishes a characterization for positive semidefiniteness of matrices from $\mathbf{S}^{(v)}$, but it does not give the explicit formulas for the matrices $B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k}$. We next develop an equivalent reformulation of positive semidefiniteness of matrices from $\mathbf{S}^{(v)}$ by introducing some additional variables.

Lemma 3.3.2 *Let $m > 1$. A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if there exists a matrix $\Delta^{m-1} = (\Delta^{m-1})^T = [\Delta^{m-1}_{ij}]_{i,j\in J'_m}$ such that the matrices*

$$\mathcal{B} \equiv \mathcal{B}_m(A, \Delta^{m-1}) = [B_{ij}]_{i,j\in J_m} : B_{ij} = \begin{cases} A_{ij}, & i \notin J'_m \text{ or } j \notin J'_m \\ \Delta^{m-1}_{ij}, & i, j \in J'_m \end{cases} \tag{107}$$

and

$$\mathcal{C} \equiv \mathcal{C}_m(A, \Delta^{m-1}) = [C_{ij}]^{i_{m-1}}_{i,j=1} : C_{ij} = \begin{cases} A_{ij}, & i \notin J'_m \text{ or } j \notin J'_m \\ A_{ij} - \Delta^{m-1}_{ij}, & i, j \in J'_m \end{cases} \tag{108}$$

are positive semidefinite.

Proof. A is the sum of matrices obtained from $\mathcal{B}$ and $\mathcal{C}$ by adding a number of zero rows and columns; thus, if $\mathcal{B}$ and $\mathcal{C}$ are $\succeq 0$, so is A. Conversely, assuming $A \succeq 0$, let us prove that there exists Δ^{m-1} such that the corresponding matrices $\mathcal{B}$, $\mathcal{C}$ are $\succeq 0$. Let B^m be defined in (106). Recall from the proof of Proposition 3.3.1(i) that $B^m \succeq 0$ and $A - B^m \succeq 0$. Now, let $\Delta^{m-1} = \left[B^m_{ij}\right]_{i,j\in J'_m}$. From the construction of B^m, we see that $\mathcal{B}$ defined as in (107) satisfies $\mathcal{B} = \left[B^m_{ij}\right]_{i,j\in J_m}$, and hence $\mathcal{B} \succeq 0$. Similarly, $\mathcal{C}$ defined as in (108) is actually the North-Western $i_{m-1} \times i_{m-1}$ block in $A - B^m$, and hence $\mathcal{C} \succeq 0$. ∎

Observing that matrix $\mathcal{C} = \mathcal{C}_m(A, \Delta^{m-1})$ belongs to $\mathbf{S}^{(v')}$, where $v' = (v'_1, ..., v'_{i_{m-1}})^T$, where $v'_i = \min[v_i, i_{m-1} - i]$, $1 \le i \le i_{m-1}$, and applying Lemma 3.3.2 recursively, we arrive at the following result.

Theorem 3.3.3 *Let $v \in \mathbf{R}^n$ be an integral nonnegative vector such that $i + v_i \le n$ for all i, let $I = \{i_1 < i_2 < ... < i_m\}$ be the image of $\{1, 2, ..., n\}$ under the mapping $i \mapsto i + v_i$, and let the sets J_k, J'_k be defined by (101). A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if this matrix can be extended, by properly chosen matrices $\Delta^k = [\Delta_k]^T = \left[\Delta^k_{ij}\right]_{i,j\in J'_{k+1}}$, $k = 1, 2, ..., m-1$, to*

a solution of an explicit system $\mathcal{S}$ of m LMIs

$$\mathcal{B}_k(A,\Delta)\succeq 0,\ k=1,...,m$$

given by the following recurrence:

Initialization: *Set $k=m$, $\mathcal{C}^m=A$.* Step k, $m\geq k\geq 1$: *Given matrix $\mathcal{C}^k\in\mathbf{S}^{(v^k)}$, with $v_i^k=\min[i_k-i,v_i]$, $i=1,2,\ldots,i_k$, set*

$$\mathcal{B}_k(\mathcal{A},\Delta)=[B_{ij}^k]_{i,j\in J_k}:B_{ij}^k=\begin{cases}\mathcal{C}_{ij}^k, & i\in J_k\backslash J_k' \text{ or } j\in J_k\backslash J_k'\\ \Delta_{ij}^{k-1}, & i,j\in J_k'\end{cases}$$

If $k=1$, terminate, otherwise set

$$\mathcal{C}^{k-1}=\left[C_{ij}^{k-1}\right]_{i,j=1}^{i_{k-1}}:C_{ij}^{k-1}=\begin{cases}\mathcal{C}_{ij}^k, & i\notin J_k' \text{ or } j\notin J_k'\\ \mathcal{C}_{ij}^k-\Delta_{ij}^{k-1}, & i,j\in J_k'\end{cases}$$

replace k with $k-1$ and loop.

From the construction of $\mathcal{B}_k\equiv\mathcal{B}_k(A,\Delta)$ above, we see that each cell ij with $i\leq j$ belongs to $\mathcal{B}_k$ exactly for all $k\in D_{ij}$ and for those k the corresponding entry ij in B^k is

$$B_{ij}^k=\begin{cases}A_{ij}, & k_-(i,j)=k=k_+(i,j)\\ A_{ij}-\sum\limits_{\nu=k_-(i,j)}^{k_+(i,j)-1}\Delta_{ij}^{\nu}, & k_-(i,j)=k<k_+(i,j)\\ \Delta_{ij}^{k-1}, & k_-(i,j)<k\leq k_+(i,j)\end{cases}\qquad(109)$$

Note that Δ^k is the principal sub-matrix in $\mathcal{B}_{k+1}$ corresponding to $i,j\in J_{k+1}'$, and that A is the sum of matrices obtained from $\mathcal{B}_1,\ldots,\mathcal{B}_m$ by adding zero rows and columns. We arrive at the following result.

Theorem 3.3.4 *A matrix $A\in\mathbf{S}^{(v)}$ is $\succeq 0$ if and only if there exist matrices $\Delta^k=[\Delta^k]^T=[\Delta_{ij}^k]_{i,j\in J_{k+1}'}$, $1\leq k\leq m-1$, such that the matrices $\mathcal{B}_k=\mathcal{B}_k(A,\{\Delta^k\}_{k=1}^{m-1})=[B_{ij}^k]_{i,j\in J_k}$ given by* (109) *are $\succeq 0$. Whenever this is the case, one has*

$$\begin{aligned}\Delta^k &\succeq 0,\ k=1,...,m-1\\ \sum_{k=1}^{m-1}\mathrm{Tr}(\Delta^k) &\leq \mathrm{Tr}(A).\end{aligned}$$

Let $v \in \mathbf{R}^n$ be a simple sparsity structure, and J_k, J'_k, $k = 1, ..., m$, be the corresponding index sets (see Section 3.2). We define $\mathbf{\Delta}$ as an Euclidean space comprised of collections $\Delta = \{\Delta^k = [\Delta^k]^T = [\Delta^k_{ij}]_{i,j \in J'_{k+1}}\}_{k=1}^{m-1}$, i.e.,

$$\mathbf{\Delta} \equiv \left\{\Delta = \{\Delta^k = [\Delta^k]^T = [\Delta^k_{ij}]_{i,j \in J'_{k+1}}\}_{k=1}^{m-1}\right\}.$$

Let $\mathbf{\Delta}_\rho$ be a subset of $\mathbf{\Delta}$ defined as

$$\mathbf{\Delta}_\rho = \left\{\Delta \in \mathbf{\Delta} : \Delta^k \succeq 0,\, k = 1, ..., m-1,\, \sum_{k=1}^{m-1} \mathrm{Tr}(\Delta^k) \leq \rho\right\}. \tag{110}$$

We denote by $\mathcal{B}_k(A, \Delta) = [B^k_{ij}(A, \Delta)]_{i,j \in J_k}$ the linear matrix-valued functions of $A \in \mathbf{S}^{(v)}$, $\Delta \in \mathbf{\Delta}$ defined by (109). Finally, let

$$\lambda_{\min}(A, \Delta) = \min_{1 \leq k \leq m} \lambda_{\min}(\mathcal{B}_k(A, \Delta)).$$

The following proposition will be used in Section 3.4.

Proposition 3.3.5 *Let $A \in \mathbf{S}^{(v)}$, $\Delta \in \mathbf{\Delta}$ be such that $\lambda_{\min}(A, \Delta) = -\lambda < 0$. Then $A \succeq -\lambda\mathcal{K}$, where $\mathcal{K}$ is given by* (102).

Proof. Let $\widehat{\Delta}^k_{ij} = \begin{cases} \Delta^k_{ij}, & i \neq j \\ \Delta^k_{ij} + \lambda, & i = j \end{cases}$, and let $\widehat{A} = A + \lambda\mathcal{K}$. By (109), we have

$$i, j \in J_k \Rightarrow B^k_{ij}(\widehat{A}, \widehat{\Delta}) - B^k_{ij}(A, \Delta) = \lambda\delta^i_j,$$

whence $\mathcal{B}_k(\widehat{A}, \widehat{\Delta}) \succeq 0$, and $\widehat{A} = A + \lambda\mathcal{K} \succeq 0$. ∎

Sizes of $\mathcal{S}$. We have expressed positive semidefiniteness of $A \in \mathbf{S}^{(v)}$ as solvability of an explicit system $\mathcal{S}$ (see Theorem 3.3.3) of LMIs in matrix A and additional matrix variables Δ^k, $k = 1, ..., m-1$. The sizes of $\mathcal{S}$ are as follows:

1. *Number and sizes of LMIs.* $\mathcal{S}$ contains m LMIs $\mathcal{B}_k(A, \Delta) \succeq 0$ of row sizes $S_k = |J_k|$, $k = 1, ..., m$.

2. *Number of additional variables.* Let $d_k = i_k - i_{k-1}$, $k = 1, ..., m$. Clearly, step $k \geq 2$ of our construction adds $V_k = \frac{(|J_k| - d_k)(|J_k| - d_k + 1)}{2}$ additional variables, and step $k = 1$ does not add new variables. Thus, the total number of additional variables is

$$V = \sum_{k=2}^{m} \frac{(|J_k| - d_k)(|J_k| - d_k + 1)}{2}.$$

Example: staircase structure. Before ending this subsection, we present an example for positive semidefinite *staircase* matrices to illustrate the result established in Theorem 3.3.4.

Let $d = (d_0, d_1, ..., d_\mu)$ be a *staircase structure* - collection of integers with $d_0 \geq 0$ and $d_1, ..., d_\mu > 0$, and let $|d| = d_0 + ... + d_\mu$. Let $\mathbf{S}^{[d]}$ be the subspace of *d-staircase* symmetric matrices in $\mathbf{S}^{|d|}$, which is comprised of $(\mu+1)\times(\mu+1)$ block matrices $[A_{ij}]_{i,j=0}^{\mu}$ with $d_i \times d_j$ blocks A_{ij} such that $A = A^T$ and $A_{ij} = 0$ for $0 < i < j-1$:

$$A \in \mathbf{S}^{[d]} \Leftrightarrow A = \left[\begin{array}{c|c|c|c|c|c} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,\mu-1} & A_{0,\mu} \\ \hline A_{0,1}^T & A_{1,1} & A_{1,2} & 0 & 0 & 0 \\ \hline A_{0,2}^T & A_{1,2}^T & A_{2,2} & A_{2,3} & 0 & 0 \\ \hline \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \hline A_{0,\mu-1}^T & 0 & 0 & A_{\mu-2,\mu-1}^T & A_{\mu-1,\mu-1} & A_{\mu-1,\mu} \\ \hline A_{0,\mu}^T & 0 & 0 & 0 & A_{\mu-1,\mu}^T & A_{\mu,\mu} \end{array}\right].$$

In view of the definition of simple sparsity structure, we easily see that $A \in \mathbf{S}^{[d]}$ iff $A \in \mathbf{S}^{(v)}$, where v is a simple sparsity structure defined as

$$v_i = \begin{cases} |d| - i, & i \leq d_0 \\ \sum_{j=0}^{k+1} d_j - i, & \sum_{j=0}^{k-1} d_j < i \leq \sum_{j=0}^{k} d_j \text{ for } k = 1, \cdots, \mu-1 \\ |d| - i, & \sum_{j=0}^{\mu-1} d_j < i \leq |d| \end{cases}$$

We also see that there are $m = \mu - 1$ elements $i_1 < ... < i_m$ in I given by $i_k = \sum_{j=0}^{k+1} d_j$ for $k = 1, \cdots, m$. Using Theorem 3.3.4, we immediately have the following result.

Proposition 3.3.6 *A d-staircase matrix $A = [A_{ij}]_{i,j=0}^{\mu}$ is positive semidefinite if and only if there exists*

$$\Delta = \left\{ \Delta^j = \left[\begin{array}{c|c} \Delta_{0,0}^j & \Delta_{0,1}^j \\ \hline [\Delta_{0,1}^j]^T & \Delta_{1,1}^j \end{array}\right] : \Delta_{0,0}^j \in \mathbf{S}^{d_0}, \Delta_{1,1}^{j+1} \in \mathbf{S}^{d_j} \right\}_{j=1}^{\mu-2}$$

such that

$$\left[\begin{array}{c|c|c} A_{0,0}-\sum\limits_{j=1}^{\mu-2}\Delta_{0,0}^{j} & A_{0,1} & A_{0,2}-\Delta_{0,1}^{1} \\ \hline A_{0,1}^{T} & A_{1,1} & A_{1,2} \\ \hline A_{0,2}^{T}-[\Delta_{0,1}^{1}]^{T} & A_{1,2}^{T} & A_{2,2}-\Delta_{2,2}^{1} \end{array}\right] \succeq 0,$$

$$\left[\begin{array}{c|c|c} \Delta_{0,0}^{j-1} & \Delta_{0,1}^{j-1} & A_{0,j+1}-\Delta_{0,1}^{j} \\ \hline [\Delta_{0,1}^{j-1}]^{T} & \Delta_{1,1}^{j-1} & A_{j,j+1} \\ \hline A_{0,j+1}^{T}-[\Delta_{0,1}^{j}]^{T} & A_{j,j+1}^{T} & A_{j+1,j+1}-\Delta_{1,1}^{j} \end{array}\right] \succeq 0,\ j=2,...,\mu-2$$

$$\left[\begin{array}{c|c|c} \Delta_{0,0}^{\mu-2} & \Delta_{0,1}^{\mu-2} & A_{0,\mu} \\ \hline [\Delta_{0,1}^{\mu-2}]^{T} & \Delta_{1,1}^{\mu-2} & A_{\mu-1,\mu} \\ \hline A_{0,\mu}^{T} & A_{\mu-1,\mu}^{T} & A_{\mu,\mu} \end{array}\right] \succeq 0.$$

3.3.2 Positive semidefinite completion of matrices from $\mathbf{S}^{(v)}$

Let $\mathbf{S}_{+}^{(v)}=\mathbf{S}^{(v)}\cap\mathbf{S}_{+}^{n}$ and $\mathbf{C}^{(v)}=\{Z\in\mathbf{S}^{(v)}:\mathrm{Tr}(XZ)\geq 0\ \forall X\in\mathbf{S}_{+}^{(v)}\}$. Since the subspace $\mathbf{S}^{(v)}$ clearly intersects the interior of $\mathbf{S}_{+}^{n}$, the cone $\mathbf{C}^{(v)}$ is exactly the cone of matrices Z from $\mathbf{S}^{(v)}$ admitting positive semidefinite completion, that is, those Z which can be made positive semidefinite by replacing "hard zero" entries ij (those with $j>i+v_i$ or $i>j+v_j$) with appropriate, perhaps nonzero, entries. Proposition 3.3.1 implies the following result.

Proposition 3.3.7 *A matrix $Z=[Z_{ij}]_{i,j=1}^{n}\in\mathbf{S}^{(v)}$ belongs to $\mathbf{C}^{(v)}$ if and only if all matrices $[Z_{ij}]_{i,j\in J_k}$, $k=1,2,...,m$, are $\succeq 0$.*

Proof. By Proposition 3.3.1, $Z\in\mathbf{C}^{(v)}$ if and only if the optimal value in the optimization problem

$$\min_{\{]X_{ij}^{k}[_{i,j\in J_k}\succeq 0\}_{k=1}^{m}}\left\{\mathrm{Tr}\left(Z\sum_{k=1}^{m}]X_{ij}^{k}[_{i,j\in J_k}\right)\right\}$$

is ≥ 0. Clearly, this is so if and only if

$$\min_{]X_{ij}^{k}[_{i,j\in J_k}\succeq 0}\left\{\mathrm{Tr}\left(]Z_{ij}^{k}[_{i,j\in J_k}]X_{ij}^{k}[_{i,j\in J_k}\right)\right\}\geq 0,\ k=1,\ldots,m,$$

which implies that $]Z^k_{ij}[_{i,j\in J_k} \succeq 0$ for $k = 1, \ldots, m$ due to a well-known result that, for a real symmetric matrix A, $\min_{X\succeq 0} \mathrm{Tr}(AX) \geq 0$ if and only if $A \succeq 0$. In other words, $Z \in \mathbf{C}^{(v)}$ if and only if $[Z_{ij}]_{i,j\in J_k} \succeq 0$, $k = 1, ..., m$. ∎

Remark 3.3.8 *The result stated in Proposition 3.3.7 can be shown, with a moderate effort, as a particular case of the results of [46] on necessary and sufficient conditions for a partially defined symmetric matrix to admit positive semidefinite completion.* ∎

Corollary 3.3.9 *For $A \in \mathbf{S}^{(v)}$ one has*

$$\lambda_{\max}(A) = \max_Y \left\{ \mathrm{Tr}(AY) : Y \in \mathbf{S}^{(v)}, \mathrm{Tr}(Y) = 1, [Y_{ij}]_{i,j\in J_k} \succeq 0, k = 1, 2, ..., m \right\}. \qquad (111)$$

Indeed, for $A \in \mathbf{S}^n$ we have $\lambda_{\max}(A) = \max_Y \{\mathrm{Tr}(AY) : Y \in \mathbf{S}^n_+, \mathrm{Tr}(Y) = 1\}$; when $A \in \mathbf{S}^{(v)}$, the latter formula clearly can be rewritten as

$$\lambda_{\max}(A) = \max_Y \left\{ \mathrm{Tr}(AY) : Y \in \mathbf{C}^{(v)}, \mathrm{Tr}(Y) = 1 \right\}.$$

Invoking Proposition 3.3.7, we arrive at (111).

Before ending this subsection, we give an example on positive semidefinite completion of staircase matrices from $\mathbf{S}^{[d]}$ to illustrate the result established in Proposition 3.3.7.

Proposition 3.3.10 *Let d be a staircase structure with $\mu > 1$, and $\mathbf{C}^{[d]}$ be the cone of d-staircase matrices B admitting positive semidefinite completion. Then, a matrix $B \in \mathbf{S}^{[d]}$ belongs to $\mathbf{C}^{[d]}$ if and only if*

$$\left[\begin{array}{c|c|c} B_{0,0} & B_{0,j} & B_{0,j+1} \\ \hline B^T_{0,j} & B_{j,j} & B_{j,j+1} \\ \hline B^T_{0,j+1} & B^T_{j,j+1} & B_{j+1,j+1} \end{array}\right] \succeq 0,\ j = 1, ..., \mu - 1.$$

3.4 Using the representations

In this section, we will use the representations presented in Subsections 3.3.1 and 3.3.2 to reformulate some large-scale SDP problems into saddle point problems. The saddle point problem reformulations for a class of SDPs, and SDP relaxations of Lovász capacity and MAXCUT problems are given in Subsections 3.4.1, 3.4.2 and 3.4.3, respectively.

3.4.1 Semidefinite programs with well-structured sparse constraint matrices

Let v be a simple sparsity pattern. Consider the semidefinite program

$$\mathrm{Opt} = \max_x \left\{ c^T x : x \in X, A[x] \succeq 0 \right\}, \tag{112}$$

where X is a "simple" (see below) convex compact set in $\mathbf{R}^N$ and $A[x]$ is affine matrix-valued function on X taking values in $\mathbf{S}^{(v)}$.

Throughout this subsection, we make the following assumptions:

A.1. *We know a point $\bar{x} \in X$ such that $A[\bar{x}] \succ 0$;*

A.2. *We are given a finite upper bound,* Opt^{up}, *on the optimal value* Opt *in* (112)*;*

A.3. *We are given a finite upper bound, ρ, on the quantity*

$$\max_x \{\mathrm{Tr}(A[x]) : x \in X, A[x] \succeq 0\}.$$

Given a point $\bar{x}$ mentioned in A.1, let

$$\nu = \max\{t : A[\bar{x}] \succeq t\mathcal{K}\}, \tag{113}$$

where $\mathcal{K}$ is defined in (102). We start with the following simple fact (a kind of "exact penalty" statement):

Lemma 3.4.1 *Let $\mathcal{Y} = \{Y = \{Y^k = [Y^k_{ij}]_{i,j\in J_k}\}_{k=1}^m : Y^k \succeq 0, \sum\limits_k \mathrm{Tr}(Y^k) \leq 1\}$. Given $T \geq 0$, let us associate with* (112) *the saddle point problem*

$$\begin{array}{c} \max\limits_{x\in X, \Delta\in\mathbf{\Delta}_\rho} F_T(x,\Delta) \\ F_T(x,\Delta) = \min\limits_{Y\in\mathcal{Y}} \left[c^T x + T \sum\limits_{k=1}^m \mathrm{Tr}(Y^k \mathcal{B}_k(A[x],\Delta)) \right] \end{array} \tag{114}$$

(for the definition of $\mathbf{\Delta}_\rho$, see (110)*). Assume that*

$$T \geq \frac{1}{\nu}(\mathrm{Opt} - c^T\bar{x}). \tag{115}$$

Let $(x_\epsilon, \Delta_\epsilon)$ be an ϵ-solution to (114)*, that is, $x_\epsilon \in X$, $\Delta_\epsilon \in \mathbf{\Delta}_\rho$ and $F_T(x_\epsilon,\Delta_\epsilon) \geq \max\limits_{x\in X,\Delta\in\mathbf{\Delta}_\rho} F_T(x,\Delta) - \epsilon$. Then, the point*

$$x^\epsilon = \frac{1}{1+\gamma}x_\epsilon + \frac{\gamma}{1+\gamma}\bar{x},\ \gamma = \frac{\max[0, -\lambda_{\min}(A[x_\epsilon],\Delta_\epsilon)]}{\nu},$$

is a feasible ϵ-solution to (112)*, i.e., $x^\epsilon \in X$, $A[x^\epsilon] \succeq 0$, and $c^T x^\epsilon \geq \mathrm{Opt} - \epsilon$.*

Proof. We clearly have

$$F_T(x,\Delta) = c^T x + T\min[\lambda_{\min}(A[x],\Delta),0].$$

Further, by Theorem 3.3.4, $A[x]$ with $x \in X$ is $\succeq 0$ if and only if $\max\limits_{\Delta\in\mathbf{\Delta}_\rho} \lambda_{\min}(A[x],\Delta) \geq 0$; thus, when x is feasible for (112), we have $\sup\limits_{\Delta\in\mathbf{\Delta}_\rho} F_T(x,\Delta) \geq c^T x$, so that the optimal value of (114) is $\geq$ Opt. Consequently, ϵ-optimality of x_ϵ for (114) implies that

$$F_T(x_\epsilon,\Delta_\epsilon) \equiv c^T x_\epsilon + T\min[\lambda_{\min}(A[x_\epsilon],\Delta_\epsilon),0] \geq \text{Opt} - \epsilon. \tag{116}$$

It is possible that $\lambda_{\min}(A[x_\epsilon],\Delta_\epsilon) \geq 0$; then x_ϵ is feasible for (112) by Theorem 3.3.4, $x^\epsilon = x_\epsilon$, and (116) says that x^ϵ is a feasible ϵ-solution to (112). Now let $\lambda_{\min}(A[x_\epsilon],\Delta_\epsilon) = -\lambda < 0$, so that $\gamma = \lambda/\nu$. Then (116) implies that

$$c^T x_\epsilon + \gamma c^T \bar{x} \geq \text{Opt} - \epsilon + T\lambda + \gamma c^T \bar{x} \geq \text{Opt}(1+\gamma) - \epsilon$$

where we have used (115) to get $T\lambda \geq \gamma(\text{Opt} - c^T\bar{x})$, whence $c^T x^\epsilon \geq \text{Opt} - \epsilon$. It remains to note that $A[x_\epsilon] \succeq -\lambda\mathcal{K}$ by Proposition 3.3.5, while $A[\bar{x}] \succeq \nu\mathcal{K}$; it follows that

$$A[x^\epsilon] = (1+\gamma)^{-1}(A[x_\epsilon] + \gamma A[\bar{x}]) \succeq (1+\gamma)^{-1}\left[-\lambda\mathcal{K} + \frac{\lambda}{\nu}\nu\mathcal{K}\right] = 0.$$

■

Lemma 3.4.1 combines with the results of [90] to yield the following

Theorem 3.4.2 *Consider problem* (112) *satisfying Assumptions A.1 – A.3, and let X be either*

(a) *the Euclidean ball $\{x \in \mathbf{R}^N : \|x\|_2 \leq R\}$, or the intersection of this ball with nonnegative orthant,*

or

(b) *the box $\{x \in \mathbf{R}^N : \|x\|_\infty \leq R\}$,*

or

(c) *the $\|\cdot\|_1$-ball $\{x \in \mathbf{R}^N : \|x\|_1 \leq R\}$, or the full-dimentional simplex $\{x \in \mathbf{R}^N : 0 \leq x, \sum\limits_i x_i \leq R\}$, or the "flat" simplex $\{x \in \mathbf{R}^N : 0 \leq x, \sum\limits_i x_i = R\}$.*

Assume that we are given an upper bound χ on the norm of the homogeneous part of $A[\cdot]$ considered as a linear mapping from $(\mathbf{R}^N, \|\cdot\|_X)$ to $(\mathbf{S}^{(v)}, \|\cdot\|)$, where $\|\cdot\|_X$ is $\|\cdot\|_2$ in

the cases of (a), (b), *and is* $\|\cdot\|_1$ *in the case of* (c), *while* $\|\cdot\|$ *is the standard matrix norm (the largest singular value) throughout the remaining part of this subsection.*

Under the outlined assumptions, for every $\epsilon > 0$ *one can find a feasible* ϵ*-solution* x_ϵ *to* (112) *(so that* $x^\epsilon \in X$, $A[x^\epsilon] \succeq 0$ *and* $c^T x^\epsilon \le \mathrm{Opt} + \epsilon$*) in no more than*

$$N(\epsilon) = O(1)\frac{[\mathrm{Opt}^{\mathrm{up}} - c^T\bar{x}]\sqrt{\ln n}}{\nu\epsilon} \times \begin{cases} [\chi R + \rho\sqrt{\ln n}], & \text{case of (a)} \\ [\chi R\sqrt{N} + \rho\sqrt{\ln n}], & \text{case of (b)} \\ [\chi R\sqrt{\ln(N)} + \rho\sqrt{\ln n}], & \text{case of (c)} \end{cases}, \tag{117}$$

steps, with computational effort per step dominated by the necessity

- *to compute* $A[x]$*, for a given* x*;*
- *to compute, given* m *symmetric matrices of the row sizes* $|J_k|$, $k = 1, ..., m$, *the eigenvalue decompositions of the matrices.*

Above, $O(1)$ *is an absolute constant,* $N = \dim x$, n *is the row dimension of* $A[\cdot]$, *and* ν *is given by* (113).

Proof. Let $T = \frac{\mathrm{Opt}^{\mathrm{up}} - c^T\bar{x}}{\nu}$. By Lemma 3.4.1, a feasible ϵ-solution to (112) is readily given by an ϵ-solution to the saddle point problem (114) with T we have just defined. Now, problem (114) is of the form

$$\max_{u=(x,\Delta)\in X\times\mathbf{\Delta}_\rho} \min_{Y\in\mathcal{Y}\subset\mathbf{S}} \left[\mathrm{lin}(u, Y) + T\langle\mathcal{A}(x) + \mathcal{D}(\Delta), Y\rangle\right], \tag{118}$$

where

- $\mathrm{lin}(u, Y)$ is an appropriate affine function of u, Y,
- $Y = \mathrm{Diag}\{Y^1, ..., Y^m\}$, $Y^k = [Y^k_{ij}]_{i,j\in J_k}$, $k = 1, ..., m$, $\mathbf{S}$ is the linear space of all block-diagonal matrices Y of the indicated block-diagonal structure, and $\mathcal{Y} = \{Y \in \mathbf{S} : 0 \preceq Y, \mathrm{Tr}(Y) \le 1\}$;
- $\mathcal{A}(\cdot)$ is the linear mapping from $\mathbf{R}^N$ into $\mathbf{S}$ defined as follows. Given $x \in \mathbf{R}^N$, we compute the homogeneous part $A = A(x) = A[x] - A[0]$ of the mapping $A[\cdot]$ at x. The k-th diagonal block $\mathcal{A}^k(x)$ in $\mathcal{A}(x)$, $k = 1, ..., m$, is the contribution of A to $\mathcal{B}_k(A[x], \Delta)$, see (109);
- $\mathcal{D}(\cdot)$ is the linear mapping from the space $\widehat{\mathbf{S}}$ of block-diagonal matrices $\Delta = \mathrm{Diag}\{\Delta^1, ..., \Delta^{m-1}\}$, $\Delta^\ell = [\Delta^\ell_{ij}]_{i,j\in J'_{\ell+1}}$, $\ell = 1, \ldots, m-1$, into $\mathbf{S}$ defined as follows: The k-th diagonal block $\mathcal{D}^k(\Delta)$ in $\mathcal{D}(\Delta)$ is the contribution of Δ to $\mathcal{B}_k(A[x], \Delta)$, see (109);

• $\mathbf{\Delta}_\rho$ is the set of all positive semidefinite matrices from $\widehat{\mathbf{S}}$ with trace $\le \rho$;

• finally, $\langle \cdot, \cdot \rangle$ is the Frobenius inner product on $\mathbf{S}$.

Now, as shown in [90], the Mirror-Prox algorithm from [90] solves problem (118) within any given accuracy $\epsilon > 0$ in no more than

$$N(\epsilon) = O(1)T\frac{L_{XY}\sqrt{\Theta_X\Theta_Y} + L_{\Delta Y}\sqrt{\Theta_\Delta\Theta_Y}}{\epsilon}$$

steps of the complexity indicated in Theorem 3.4.2, where

$$\Theta_X = \begin{cases} R^2, & \text{case (a)} \\ R^2 N, & \text{case (b)} \\ R^2 \ln N, & \text{case (c)} \end{cases}, \ \Theta_Y = \ln n, \ \Theta_\Delta = \rho^2 \ln n,$$

L_{XY} is the norm of the linear mapping $\mathcal{A}$ considered as a mapping from $(\mathbf{R}^N, \|\cdot\|_X)$ to $(\mathbf{S}, \|\cdot\|)$, and $L_{\Delta Y}$ is the norm of the linear mapping $\mathcal{D}$ considered as the mapping from $(\widehat{\mathbf{S}}, |\cdot|_1)$ to $(\mathbf{S}, \|\cdot\|)$, where $|\Delta|_1$ is the sum of modulae of eigenvalues of $\Delta \in \widehat{\mathbf{S}}$. It remains to evaluate L_{XY} and $L_{\Delta Y}$. Let $x \in \mathbf{R}^N$ satisfy $\|x\|_X \le 1$, and let $A = A(x)$, so that $\|A\| \le \chi$. Invoking (109), it is immediately seen that $\mathcal{A}^k(x)$, for every k, is a "border" in A: there exist two principal submatrices in A embedded one into another such that $\mathcal{A}^k(x)$ is obtained from the larger submatrix by replacing the entries belonging to the smaller one by zeros. By Eigenvalue Interlacement Theorem, both submatrices are of norm $\le \chi$, so that the "border" is of norm $\le 2\chi$, whence $\|\mathcal{A}(x)\| \le 2\chi$. Thus, $L_{XY} \le 2\chi$. Now let us bound $L_{\Delta Y}$. The extreme points of the unit $|\cdot|_1$-ball D in $\widehat{\mathbf{S}}$ are block-diagonal matrices with just one nonzero diagonal block, which is a symmetric rank 1 matrix of the corresponding size with the only nonzero singular value equal to 1, or equivalently, is a rank 1 matrix of the Frobenius norm equal to 1. For such a matrix Δ, it follows immediately from (109) that the Frobenius (and then – the matrix) norm of every block in $\mathcal{D}(\Delta)$ is at most 2. Since $L_{\Delta Y}$ is the maximum of the quantities $\|\mathcal{D}(\Delta)\|$ over the extreme points Δ of D, we conclude that $L_{\Delta Y} \le 2$. Combining our observations, we arrive at (117). ∎

We have presented a rather general approach to solving SDPs by reducing them to saddle point problems which are further solved by the $O(t^{-1})$-converging Mirror-Prox algorithm from [90]. In the sequel, we apply this scheme to the problems of computing Lovász capacity

of a graph and to MAXCUT, with emphasis on utilizing favourable sparsity patterns of the underlying graphs.

3.4.2 Computing Lovász capacity for a graph with a favourable sparsity pattern

Let $v = (v_1, v_2, ..., v_{n+1})^T \in \mathbf{R}^{n+1}$ be a simple sparsity structure with $v_1 = n$, and let G be an undirected graph with n nodes, indexed by $2, 3, ..., n+1$, and the set of arcs E such that if $(i, j) \in E$ and $i \leq j$, then $j \leq i + v_i$. Recall from Section 3.2 that, for each entry ij with $1 \leq i \leq j \leq i + v_i$, $\ell(i, j)$ is the number of sets J_k $(k = 1, ..., m)$ such that $i, j \in J_k$. Let

$$\eta = \max_{2 \leq i \leq n+1} \ell(i, i). \tag{119}$$

Consider the Lovász capacity problem

$$\begin{array}{rcl} \vartheta(G) & = & \min\limits_{X,\lambda} \left\{ \lambda : \lambda I_n - ee^T - X \succeq 0, (i,j) \notin E \Rightarrow X_{i-1,j-1} = 0 \right\} \\ & = & \min\limits_{X,\lambda} \left\{ \lambda : \left[\begin{array}{c|c} \nu & \sqrt{\nu} e^T \\ \hline \sqrt{\nu} e & \lambda I_n - X \end{array} \right] \succeq 0, (i,j) \notin E \Rightarrow X_{i-1,j-1} = 0 \right\} \end{array} \tag{120}$$

where $e \in \mathbf{R}^n$ is the vector of ones and $\nu > 0$ is a parameter. Note that the equivalence of the two optimization problems in (120) is given by the Schur Complement Lemma. Let $\mathcal{M}$ be the affine subspace in $\mathbf{S}^{(v)}$ comprised of all matrices of the form $\left[\begin{array}{c|c} \nu & \sqrt{\nu} e^T \\ \hline \sqrt{\nu} e & Z \end{array} \right]$ with Z constrained by the requirements

$$Z_{11} = Z_{22} = ... = Z_{nn}; \, (i < j \ \& \ (i,j) \notin E) \Rightarrow Z_{i-1,j-1} = 0.$$

We equip $\mathbf{S}^{(v)}$ (and thus $\mathcal{M}$) with the Euclidean structure given by the inner product

$$\langle A, B \rangle_{\mathcal{L}} = \langle \mathcal{L}^{1/2} A \mathcal{L}^{1/2}, \mathcal{L}^{1/2} B \mathcal{L}^{1/2} \rangle,$$

where $\langle P, Q \rangle$ is the Frobenius inner product and $\mathcal{L} = \mathrm{Diag}\left\{ \{\ell^{-1/2}(i,i)\}_{i=1}^{n+1} \right\}$ (cf. (102)). The norm on $\mathbf{S}^{(v)}$ corresponding to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ will be denoted $\| \cdot \|_{\mathcal{L}}$. We denote by $\mathcal{P}$ the orthogonal projector of $\mathbf{S}^{(v)}$ onto $\mathcal{M}$, so that for any $A \in \mathbf{S}^{(v)}$ one has

$$\mathcal{P}(A) = \left[\begin{array}{c|c} \nu & \sqrt{\nu} e^T \\ \hline \sqrt{\nu} e & \gamma(A) I_n + \widehat{A} \end{array} \right],$$

where $\gamma(A) = \left(\sum_{i=2}^{n+1} \ell^{-1}(i,i)A_{ii}\right)\left(\sum_{i=2}^{n+1} \ell^{-1}(i,i)\right)^{-1}$ and the matrix $\widehat{A}$ is obtained from the South-Eastern $n \times n$ angular block of A by replacing all diagonal entries and all entries ij with $(i,j) \notin E$ with zeros.

Given an upper bound $\widehat{\theta} \leq n$ on the Lovász capacity, consider the following optimization problem:

$$\begin{array}{rcl}
\mathrm{Opt} & = & \min\limits_{B\in\mathbf{B}} \left\{ \lambda(B) + T\|\mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B))\|_{\mathcal{L}} : \begin{array}{c} B_k \succeq 0, k = 1, ..., m \\ \sum\limits_{k=1}^{m} \mathrm{Tr}(B_k^2) \leq R^2 \end{array} \right\}, \\
\mathcal{S}(B) & = & \sum\limits_{k=1}^{m} B^k,\ B^k =]B^k_{ij}[_{i,j\in J_k}, \\
\lambda(B) & = & \left(\sum\limits_{i=2}^{n+1} \ell^{-1}(i,i)(\mathcal{S}(B))_{ii}\right)\left(\sum\limits_{i=2}^{n+1} \ell^{-1}(i,i)\right)^{-1} = (\mathcal{P}(\mathcal{S}(B)))_{jj},\ j = 2, 3, ..., n+1, \\
R & = & \sqrt{\widehat{\theta}^2(n + 2|E|) + \nu^2 + 2\nu n},
\end{array} \tag{121}$$

where $\mathbf{B}$ is as defined in Subsection 3.3.1 and $T \geq 1$.

Observe that

$$\mathrm{Opt} \leq \vartheta(G). \tag{122}$$

Indeed, let X_* be the X-component of the optimal solution to (120). Then the matrix $Y_* = \left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & \vartheta(G)I_n - X_* \end{array}\right]$ is $\succeq 0$ and belongs to $\mathbf{S}^{(v)}$; by Proposition 3.3.1, this matrix is $\mathcal{S}(B^*)$ for certain $B^* \in \mathbf{B}$ with components $B^*_k \succeq 0$. From the latter fact and (104) it follows $\sum\limits_k \|B^*_k\|_F^2 \leq \|Y_*\|_F^2 \leq R^2$, with the latter inequality readily given by the fact that $|(X_*)_{ij}| \leq \vartheta(G)$ due to $Y_* \succeq 0$. Thus, B^* is feasible for (121); at this feasible solution, the objective of (121) clearly is equal to $\lambda(B^*) = \vartheta(G)$, and (122) follows.

Observe also that (121) is nothing but the saddle point problem

$$\min_{B\in\mathcal{B}} \max_{Y\in\mathcal{Y}} F(B, Y), \tag{123}$$

where

$$\begin{array}{rcl}
\mathcal{B} & = & \{B \in \mathbf{B} : B_k \succeq 0,\ k = 1, ..., m, \sum\limits_k \|B_k\|_F^2 \leq R^2\} \\
\mathcal{Y} & = & \{Y \in \mathbf{S}^{(v)} : \|Y\|_{\mathcal{L}} \leq 1\} \\
F(B, Y) & = & \lambda(B) + T\langle Y, \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B))\rangle_{\mathcal{L}}
\end{array} \tag{124}$$

Note that by (105) the norm of the linear part of the affine mapping

$$B \mapsto \mathcal{Q}(B) = \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B)),$$

treated as the mapping from the space $\mathbf{B}$ equipped with the norm $\|B\|_F = \sqrt{\sum_{k=1}^{m} \|B_k\|_F^2}$ to the space $\mathbf{S}^{(v)}$ equipped with the norm $\|\cdot\|_{\mathcal{L}}$ is ≤ 1.

Since the mapping $\mathcal{Q}$ is of norm ≤ 1, from the results of [90] the saddle point problem (123) can be solved within accuracy $\epsilon > 0$ in no more than

$$N(\epsilon) = O(1)\frac{TR}{\epsilon} \tag{125}$$

steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m symmetric matrices of the sizes $|J_1|, ..., |J_m|$. Thus, computational effort per step does not exceed

$$\mathcal{C} = O(1)\sum_{k=1}^{m} |J_k|^3. \tag{126}$$

Assume that we have found an ϵ-solution $\widetilde{B} = \{\widetilde{B}_k\}_{k=1}^m \in \mathcal{B}$ to (123), so that

$$\lambda(\widetilde{B}) + T\underbrace{\|\overbrace{\mathcal{S}(\widetilde{B}) - \mathcal{P}(\mathcal{S}(\widetilde{B}))}^{\Delta}\|_{\mathcal{L}}}_{\delta} \leq \mathrm{Opt} + \epsilon. \tag{127}$$

and $\widetilde{B}_k \succeq 0$ for all k, whence $\mathcal{S}(\widetilde{B}) \succeq 0$. Observe that

$$\mathcal{P}(\mathcal{S}(\widetilde{B})) = \left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & \lambda(\widetilde{B})I_n - X \end{array}\right], \tag{128}$$

where X is of the structure required in (120). Since $\|\Delta\|_{\mathcal{L}} \equiv \|\mathcal{L}^{1/2}\Delta\mathcal{L}^{1/2}\|_F = \delta$ (where Δ, δ are defined as in (127)), we have $\mathcal{L}^{1/2}\Delta\mathcal{L}^{1/2} \preceq \delta I_{n+1}$, whence $\Delta \preceq \delta\mathcal{L}^{-1}$. This combined with $\mathcal{S}(\widetilde{B}) \succeq 0$ results in $\mathcal{P}(\mathcal{S}(\widetilde{B})) \succeq -\delta\mathcal{L}^{-1}$. This together with (102), (119) and (128) implies that

$$\left[\begin{array}{c|c} \nu + m^{1/2}\delta & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & [\lambda(\widetilde{B}) + \eta^{1/2}\delta]I_n - X \end{array}\right] \succeq 0,$$

whence

$$\left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & \frac{\nu+m^{1/2}\delta}{\nu}[\lambda(\widetilde{B}) + \eta^{1/2}\delta]I_n - \frac{\nu+m^{1/2}\delta}{\nu}X \end{array}\right] \succeq 0.$$

Thus, an ϵ-solution $\widetilde{B}$ to (123) can be easily converted to a feasible solution $(\widetilde{\lambda}, \widetilde{X} = \frac{\nu+m^{1/2}\delta}{\nu}X)$ to (120) with the value of the objective

$$\begin{array}{rcll}
\widetilde{\lambda} & = & \frac{\nu+m^{1/2}\delta}{\nu}[\lambda(\widetilde{B}) + \eta^{1/2}\delta] & \\
& \leq & \frac{\nu+m^{1/2}\delta}{\nu}\left[\mathrm{Opt} + \epsilon - (T-\eta^{1/2})\delta\right] & [\text{see (127)}] \\
& \leq & \frac{\nu+m^{1/2}\delta}{\nu}\left[\vartheta(G) + \epsilon - (T-\eta^{1/2})\delta\right] & [\text{see (122)}] \\
& = & \vartheta(G) + \epsilon + \delta\left[\frac{m^{1/2}(\vartheta(G)+\epsilon)}{\nu} - (T-\eta^{1/2})\frac{\nu+m^{1/2}\delta}{\nu}\right]. &
\end{array}$$

We arrive at the following result:

Proposition 3.4.3 *Let η be defined as in (119), and let*

$$T \geq \eta^{1/2} + \frac{m^{1/2}(\vartheta(G)+\epsilon)}{\nu}.$$

Then an ϵ-solution to (123) *induces a feasible ϵ-solution to* (120). *The number of steps required to get such a solution can be bounded by* (125), *while the computational effort per step can be bounded by* (126).

Corollary 3.4.4 *Given an upper bound $\widehat{\theta}$ on $\vartheta(G)$, let us set*

$$\phi(\nu) = \left(\eta^{1/2} + \frac{m^{1/2}\widehat{\theta}}{\nu}\right)\sqrt{\widehat{\theta}^2(n+2|E|) + \nu^2 + 2\nu n}$$

and

$$\bar{\nu} = \mathrm{argmin}_{\nu>0}\phi(\nu), \quad \widehat{T} = \eta^{1/2} + \frac{m^{1/2}\widehat{\theta}}{\bar{\nu}}$$

With $T = \widehat{T}$, the outlined procedure allows, for every ϵ, $0 < \epsilon \leq \widehat{\theta} - \vartheta(G)$, to find a feasible ϵ-solution to (120) *in no more than*

$$N(\epsilon) = O(1)\frac{\phi(\bar{\nu})}{\epsilon}$$

steps, with the complexity of a step given by (126).

Corollary 3.4.5 *Let $|E| \geq n$. Then, setting*

$$\nu = \min\left[\widehat{\theta}\sqrt{|E|}, \widehat{\theta}^2|E|n^{-1}\right],$$

one gets

$$N(\epsilon) \leq O(1)\frac{\widehat{\theta}\sqrt{m|E|}}{\epsilon}.$$

Proof. Indeed, with ν in question, we clearly have $\sqrt{\widehat{\theta}^2(n+2|E|)+\nu^2+2\nu n} \leq O(1)\widehat{\theta}\sqrt{|E|}$. Consequently,

$$\phi(\nu) \leq O(1)\widehat{\theta}\sqrt{|E|}\Big(\underbrace{\eta^{1/2}}_{\leq m^{1/2}} + \max[\underbrace{\frac{m^{1/2}}{\sqrt{|E|}}}_{\leq 1}, \underbrace{\frac{m^{1/2}n}{\widehat{\theta}|E|}}_{\leq m^{1/2}}]\Big) \leq O(1)\widehat{\theta}\sqrt{m|E|}.$$

This together with Corollary 3.4.4 implies that the conclusion holds. ∎

Example: staircase structure. Let p, q be positive integers, and $n = p(q+1)$. Assume that the incidence matrix of the graph is from $\mathbf{S}^{[d]}$, where $d \in \mathbf{R}^{q+1}$ with $d_i = p$ for $i = 0, \cdots, q$. Then, from (120), we see that

$$i + v_i = \begin{cases} n+1, & 1 \leq i < 2+p \\ 1+(k+1)p, & 2+(k-1)p \leq i < 2+kp,\ k = 2, ..., q \\ 1+(q+1)p, & 2+pq \leq i \leq n+1 \end{cases}$$

In the preceding notation, we have $i_k = 1+(k+2)p$, $k = 1, ..., q-1$, $\eta = q-1$, $|J_k| = 3p+1$, $|E| \leq O(1)p^2q$. Thus,

$$\mathcal{C} = O(1)p^3q,\ \phi(\nu) \leq O(1)\left(q^{1/2} + \frac{q^{1/2}\widehat{\theta}}{\nu}\right)\left(\widehat{\theta}^2p^2q + \nu^2 + \nu pq\right)^{1/2}.$$

Setting $\widehat{\nu} = \widehat{\theta}pq^{1/2}$, we get

$$\phi(\widehat{\nu}) \leq O(1)q^{1/2}\left(\widehat{\theta}^2p^2q + p^2q^{3/2}\widehat{\theta}\right)^{1/2} \leq O(1)\widehat{\theta}pq\left(1 + q^{1/2}\widehat{\theta}^{-1}\right)^{1/2}.$$

Since the stability number of the corresponding graph clearly is at least $O(q)$, we have $\phi(\widehat{\nu}) \leq O(1)\widehat{\theta}pq$. Consequently, computing Lovász capacity within accuracy ϵ costs at most

$$O(1)\frac{\widehat{\theta}pq}{\epsilon} \times p^3q = O(1)\frac{\widehat{\theta}p^4q^2}{\epsilon}$$

operations. For comparison:

1. Saddle point approach, similar to the above one, as applied to computing Lovász capacity for a *general* pq-node graph G with $O(p^2q)$ arcs and $\vartheta(G) \leq \widehat{\theta}$, results in the bound $O(1)\frac{\widehat{\theta}p^4q^{7/2}\sqrt{\ln(pq)}}{\epsilon}$, see [90];

2. The arithmetic cost of a single interior point iteration in the problem of computing Lovász capacity of a *general* pq-node graph is as large as $O(1)p^6q^6$, and is at least p^6q^3 even in the case of graph possessing the structure in question.

3.4.3 The MAXCUT problem on a graph with a favourable sparsity pattern

Consider a MAXCUT-type problem

$$\mathrm{Opt} = \max_{X \in \mathbf{S}^n} \{\mathrm{Tr}(VX) : X \succeq 0, \mathrm{diag}(X) = \mathbf{e}\} \tag{129}$$

where $\mathrm{diag}(A)$ is the diagonal of a square matrix A and $\mathbf{e}$ is the vector of ones. Assume that $V \in \mathbf{S}^{(v)}$ for a given simple sparsity structure v. By Proposition 3.3.7 problem (129) is equivalent to

$$\mathrm{Opt} = \max_{X \in \mathbf{S}^{(v)}} \left\{\mathrm{Tr}(VX) : \mathrm{diag}(X) = \mathbf{e}, X^k \equiv [X_{ij}]_{i,j \in J_k} \succeq 0,\ k = 1, ..., m\right\}. \tag{130}$$

Let $\mathcal{X} = \{X \in \mathbf{S}^{(v)} : |X_{ij}| \leq 1 \forall i, j, X_{ii} = 1 \forall i\}$, $\mathcal{Y} = \{Y = \{Y^k = [Y^k_{ij}]_{i,j \in J_k}\}_{k=1}^m : Y^k \succeq 0, \sum\limits_k \mathrm{Tr}(Y^k) \leq 1\}$. Consider the saddle point problem

$$\mathrm{Opt}^+ = \max_{X \in \mathcal{X}} \Phi(X) \equiv \min_{Y \in \mathcal{Y}} \left[\mathrm{Tr}(VX) + T \sum_{k=1}^m \mathrm{Tr}(X^k Y^k)\right], \tag{131}$$

where $T > 0$ is a parameter. Observe that the optimal value in (131) is $\geq \mathrm{Opt}$. Indeed, if X_* is an optimal solution to (130), then clearly $X_* \in \mathcal{X}$, and $\Phi(X_*) = \mathrm{Tr}(VX_*)$. Now let X be an ϵ-solution to (131), so that $X \in \mathcal{X}$ and

$$\begin{aligned} \mathrm{Tr}(VX) - T\lambda &\geq \mathrm{Opt}^+ - \epsilon \geq \mathrm{Opt} - \epsilon, \\ \lambda &= \max[0, -\lambda_{\min}(X^1), ..., -\lambda_{\min}(X^m)]. \end{aligned}$$

It is possible that $\lambda = 0$, that is, X is feasible for (130); in this case, X is a feasible ϵ-solution to the latter problem. Now consider the case when $\lambda > 0$, and let $\tilde{X} = (1 + \lambda)^{-1}(X + \lambda I)$. Clearly, $\tilde{X}$ is feasible for (130). Setting $\hat{X} = X + \lambda I$, we have

$$\mathrm{Tr}(V\hat{X}) = \mathrm{Tr}(VX) + \lambda \mathrm{Tr}(V) \geq \mathrm{Opt} - \epsilon + \lambda[\mathrm{Tr}(V) + T].$$

Hence, we obtain that

$$\begin{aligned} \mathrm{Tr}(V\tilde{X}) &\geq (1 + \lambda)^{-1}[\mathrm{Opt} - \epsilon] + \lambda[\mathrm{Tr}(V) + T] \\ &\geq \mathrm{Opt} - \epsilon + (1 + \lambda)^{-1}\lambda\left[T + \mathrm{Tr}(V) - \mathrm{Opt}\right]. \end{aligned}$$

We see that *if*

$$T \geq \mathrm{Opt} - \mathrm{Tr}(V),$$

then $\tilde{X}$ is a feasible ϵ-solution to (130). This observation suggests the following scheme for solving (130): given an upper bound $\mathrm{Opt}^{\mathrm{up}}$ on Opt, we set $T = \mathrm{Opt}^{\mathrm{up}} - \mathrm{Tr}(V)$ and solve saddle point problem (131) within accuracy ϵ, and then convert, in the just presented fashion, the resulting X into a feasible ϵ-solution to (130).

By [90], generating an ϵ-solution to (131) costs $O(1)\frac{T\sqrt{\dim \mathbf{S}^{(v)}}\sqrt{\ln n}}{\epsilon}$ steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m matrices X^k, $k = 1, ..., m$, where X^k is defined as in (130) for $X \in \mathcal{X}$. We arrive at the following result:

Proposition 3.4.6 *Let an upper bound* $\mathrm{Opt}^{\mathrm{up}}$ *on the optimal value in* (129) *be given. For every $\epsilon > 0$, a feasible ϵ-solution to problem* (129) *with $V \in \mathbf{S}^{(v)}$ can be found in no more than*

$$N(\epsilon) = O(1)\frac{[\mathrm{Opt}^{\mathrm{up}} - \mathrm{Tr}(V)]\sqrt{\ln n}\sqrt{\sum_{i=1}^{n}(1+v_i)}}{\epsilon} \tag{132}$$

steps of Mirror-Prox algorithm [90], with $O(1)\sum\limits_{k=1}^{m}|J_k|^3$ operations per step.

Remark 3.4.7 *When V is a diagonal-dominated matrix: $V_{ii} \geq \sum\limits_{j\neq i}|V_{ij}|$ (as it is the case in the true MAXCUT problem), one clearly has* $\mathrm{Tr}(V) \leq \mathrm{Opt} \leq 2\mathrm{Tr}(V)$. *In this case, one can set* $\mathrm{Opt}^{\mathrm{up}} = 2\mathrm{Tr}(V)$, *thus converting* (132) *into the bound*

$$N(\epsilon) \leq O(1)\frac{\mathrm{Tr}(V)}{\epsilon}\sqrt{\ln n}\sqrt{\sum_{i=1}^{n}(1+v_i)}.$$

■

Example: staircase structure. Let m, p be positive integers, and $n = p(m+1)$. Consider the staircase structure $d = (p, ..., p) \in \mathbf{R}^{m+1}$, and assume that we are given an n-node graph G with incidence matrix belonging to $\mathbf{S}^{[d]}$. Given a matrix A of nonnegative weights of arcs in G, let $V_{ij} = \frac{1}{4}\begin{cases} -A_{ij}, & i \neq j \\ \sum\limits_j A_{ij}, & i = j \end{cases}$, so that (129) becomes the standard MAXCUT problem associated with (A, G). By Remark 3.4.7, the outlined scheme allows to solve the latter problem within any accuracy $\epsilon > 0$ at the arithmetic cost of $O(1)\frac{\mathrm{Opt}}{\epsilon}p^4m^{3/2}\sqrt{\ln(pm)}$ operations. Note that the arithmetic cost of a *single* interior point iteration as applied to the "most economical" dual reformulation of (129), is $O(1)p^3m^3$. It

follows that when a "moderate" relative accuracy ϵ/Opt, say, $\epsilon/\mathrm{Opt} = 0.01$ is sought and $m^{3/2} \gg p\sqrt{\ln(pm)}$, the Mirror-Prox algorithm as applied to the MAXCUT problem by far outperforms Interior Point techniques. The difference becomes even more significant when we compare the complexity bound for Mirror-Prox with the theoretical complexity bound of $O(1)\sqrt{pm}\ln\left(\frac{\mathrm{Opt}}{\epsilon}\right)p^3m^3$ operations for IPMs (the factor $O(1)\sqrt{pm}\ln\left(\frac{\mathrm{Opt}}{\epsilon}\right)$ is the theoretical bound on the number of IPM iterations required to get an ϵ-solution).

3.5 Numerical implementation

In this section, we present the results of numerical experiments with the Lovász capacity problem (120) and the (semidefinite relaxation of the) MAXCUT problem (129). These problems were solved by the first-order Mirror-Prox algorithm from [90] as applied to the saddle point reformulations (123), respectively, (131), of the problems.

In our experiments, the incidence matrix has staircase structure with $d = (p, ..., p) \in \mathbf{R}^{m+1}$, with dense $p \times p$ blocks allowed by the structure. Note that the number of nodes in such a graph is $n = (m+1)p$, while the number of arcs is $\frac{p^2(5m-1)-p(m+1)}{2}$. For our computations, we generated graphs with $p = 2, 3, \cdots, 6$ and n ranging from about $10,000$ to about $80,000$ (so that the number of arcs varied from about 50,000 to about 1,100,000). We terminate the computations when the relative error, as given by valid on-line inaccuracy bounds generated by the Mirror-Prox algorithm, became less than 1% for both problems. Our code is written in ANSI C. All computations are performed on Supermicro dual-2.66GHz Intel Xeon server with 2GB RAM.

Lovász capacity problem. When solving this problem according to the scheme developed in Section 3.4.2, one needs an a priori upper bound $\widehat{\theta}$ on $\vartheta(G)$. Using the well-know result that Lovász capacity number of a graph G is bounded above by the chromatic number of the complement graph, it easy to see that for the graphs we are generating one can take $\widehat{\theta} = m$, and these were the upper bounds used in our computations. The results are presented in Table 1. In the table 1, the first three columns report the sizes of our generated graphs. The fourth and the fifth columns present the valid upper, respectively, lower

Table 1: Computational result for the Lovász capacity problem

(m,p)	Nodes	Edges	LwBnd	UpBnd	Iter	CPU
(4999,2)	10000	44988	2.497e3	2.516e3	11757	$3^h55'16''$
(9999,2)	20000	89988	4.996e3	5.045e3	17238	$11^h50'58''$
(14999,2)	30000	134988	7.484e3	7.545e3	30162	$32^h58'22''$
(19999,2)	40000	179988	9.952e3	1.004e4	34833	$51^h49'3''$
(3333,3)	10002	69987	1.660e3	1.676e3	10770	$4^h22'41''$
(6666,3)	20001	139980	3.329e3	3.358e3	20097	$16^h28'4''$
(9999,3)	30000	209973	4.998e3	5.046e3	24615	$33^h5'22''$
(13333,3)	40002	279987	6.643e3	6.708e3	29154	$51^h2'55''$
(2499,4)	10000	94952	1.249e3	1.259e3	8313	$4^h11'37''$
(4999,4)	20000	189952	2.491e3	2.514e3	17412	$17^h52'14''$
(7499,4)	30000	284952	3.747e3	3.784e3	21315	$34^h10'7''$
(9999,4)	40000	379952	4.972e3	5.022e3	28737	$61^h11'53''$
(1999,5)	10000	119925	9.970e2	1.001e3	9792	$5^h43'27''$
(3999,5)	20000	239925	1.995e3	2.013e3	13041	$15^h29'2''$
(5999,5)	30000	359925	2.989e3	3.016e3	23625	$42^h46'4''$
(7999,5)	40000	479925	3.990e3	4.022e3	27381	$75^h41'56''$
(1666,6)	10002	144921	8.301e2	8.382e2	9999	$6^h56'21''$
(3333,6)	20004	289950	1.659e3	1.676e3	14205	$20^h27'42''$
(4999,6)	30000	434892	2.496e3	2.517e3	17403	$46^h12'42''$
(6666,6)	40002	579921	3.331e3	3.364e3	21621	$62^h33'58''$

bounds on $\vartheta(G)$ as reported by the Mirror-Prox algorithm. The last two columns report the number of steps and the CPU time.

Semidefinite relaxation of MAXCUT. The graphs used in our experiments have the same structure as in the case of Lovász capacity problems. The weights of the arcs were picked at random from the uniform distribution in $[1, 11]$. The results are presented in Table 2; the structure of Table 2 is identical to the one of Table 1.

Table 2: Computational results for the MAXCUT problem

(m,p)	Nodes	Edges	LwBnd	UpBnd	Iter	CPU
(4999,2)	10000	44988	1.921e5	1.940e5	2604	$51'22''$
(9999,2)	20000	89988	3.859e5	3.898e5	3711	$2^h20'40''$
(14999,2)	30000	134988	5.775e5	5.833e5	3963	$3^h50'19''$
(19999,2)	40000	179988	7.725e5	7.802e5	4137	$5^h10'46''$
(39999,2)	80000	359988	1.545e6	1.560e6	5622	$13^h53'19''$
(3333,3)	10002	69987	2.862e5	2.891e5	3447	$1^h29'56''$
(6666,3)	20001	139980	5.717e5	5.775e5	3765	$3^h14'50''$
(9999,3)	30000	209973	8.574e5	8.660e5	4536	$5^h58'23''$
(13333,3)	40002	279987	1.146e6	1.157e6	5355	$9^h4'41''$
(26666,3)	80001	559980	2.291e6	2.314e6	7260	$24^h24'51''$
(2499,4)	10000	94952	3.783e5	3.821e5	2673	$1^h21'31''$
(4999,4)	20000	189952	7.585e5	7.660e5	3531	$3^h36'46''$
(7499,4)	30000	284952	1.137e6	1.148e6	4317	$6^h49'26''$
(9999,4)	40000	379952	1.515e6	1.530e6	4773	$9^h42'54''$
(19999,4)	80000	759952	3.028e6	3.058e6	6393	$25^h53'39''$
(1999,5)	10000	119925	4.703e5	4.750e5	3012	$1^h53'49''$
(3999,5)	20000	239925	9.423e5	9.517e5	3177	$3^h53'26''$
(5999,5)	30000	359925	1.417e6	1.431e6	3741	$9^h6'1''$
(7999,5)	40000	479925	1.885e6	1.904e6	4338	$10^h32'40''$
(15999,5)	80000	959925	3.771e6	3.809e6	5508	$26^h32'50''$
(1666,6)	10002	144921	5.645e5	5.701e5	2487	$1^h44'37''$
(3333,6)	20004	289950	1.127e6	1.138e6	3153	$4^h23'17''$
(4999,6)	30000	434892	1.694e6	1.711e6	3558	$9^h42'34''$
(6666,6)	40002	579921	2.257e6	2.279e6	4263	$11^h53'27''$
(13333,6)	80004	1159950	4.514e6	4.559e6	5619	$31^h17'50''$

CHAPTER IV

AN ITERATIVE SOLVER-BASED INTERIOR-POINT ALGORITHM FOR CONVEX QP

4.1 Preliminary Remarks

In this chapter we develop an interior-point long-step primal-dual infeasible path-following (PDIPF) algorithm for convex quadratic programming (CQP) whose search directions are computed by means of an iterative linear solver. We will refer to this algorithm as an *inexact* algorithm, in the sense that the Newton system which determines the search direction will only be solved approximately at each iteration. The problem we consider is

$$\min_x \left\{ \frac{1}{2} x^T Q x + c^T x : \quad Ax = b, \ x \geq 0 \right\}, \tag{133}$$

where the data are $Q \in \Re^{n \times n}$, $A \in \Re^{m \times n}$, $b \in \Re^m$, and $c \in \Re^n$, and the decision vector is $x \in \Re^n$. We also assume that Q is positive semidefinite, and that a factorization $Q = VE^2V^T$ is explicitly given, where $V \in \Re^{n \times l}$, and E is a $l \times l$ positive diagonal matrix.

A similar algorithm for solving the special case of linear programming (LP), i.e., problem (133) with $Q = 0$, was developed by Monteiro and O'Neal in [80]. The algorithm studied in [80] is essentially the long-step PDIPF algorithm studied in [58, 132], the only difference being that the search directions are computed by means of an iterative linear solver. We refer to the iterations of the iterative linear solver as the *inner iterations* and to the ones performed by the interior-point method itself as the *outer iterations*. The main step of the algorithm studied in [58, 80, 132] is the computation of the primal-dual search direction $(\Delta x, \Delta s, \Delta y)$, whose Δy component can be found by solving a system of the form $AD^2A^T\Delta y = g$, referred to as the *normal equation*, where $g \in \Re^m$ and the positive diagonal matrix D depends on the current primal-dual iterate. In contrast to [58, 132], the algorithm studied in [80] uses an iterative linear solver to obtain an approximate solution to the normal equation. Since the condition number of the normal matrix AD^2A^T may become

excessively large on degenerate LP problems (see e.g., [66]), the maximum weight basis (MWB) preconditioner T introduced in [99, 111, 124] is used to better condition this matrix and an approximate solution of the resulting equivalent system with coefficient matrix $TAD^2A^TT^T$ is then computed. By using a result obtained in [81], which establishes that the condition number of $TAD^2A^TT^T$ is uniformly bounded by a quantity depending on A only, Monteiro and O'Neal [80] show that the number of inner iterations of the algorithm in [80] can be uniformly bounded by a constant depending on n and A.

In the case of CQP, the standard normal equation takes the form

$$A(Q + X^{-1}S)^{-1}A^T\Delta y \;=\; g, \tag{134}$$

for some vector g. When Q is not diagonal, the matrix $(Q + X^{-1}S)^{-1}$ is not diagonal, and hence the coefficient matrix of (134) does not have the form required for the result of [81] to hold. To remedy this difficulty, we develop in this chapter a new linear system, referred to as the *augmented normal equation* (ANE), to determine a portion of the primal-dual search direction. This equation has the form $\tilde{A}\tilde{D}^2\tilde{A}^Tu = w$, where $w \in \Re^{m+l}$, $\tilde{D}$ is an $(n+l)\times(n+l)$ positive diagonal matrix and $\tilde{A}$ is a 2×2 block matrix of dimension $(m+l) \times (n+l)$ whose blocks consist of A, V^T, the zero matrix and the identity matrix (see equation (153)). As was done in [80], a MWB preconditioner $\tilde{T}$ for the ANE is computed and an approximate solution of the resulting preconditioned equation with coefficient matrix $\tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T$ is generated using an iterative linear solver. Using the result of [81], which claims that the condition number of $\tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T$ is uniformly bounded regardless of $\tilde{D}$, we obtain a uniform bound (depending only on $\tilde{A}$) on the number of inner iterations performed by the iterative linear solver to find a desirable approximate solution to the ANE (see Theorem 4.3.5).

Since the iterative linear solver can only generate an approximate solution to the ANE, it is clear that not all equations of the Newton system, which determines the primal-dual search direction, can be satisfied simultaneously. In the context of LP, Monteiro and O'Neal [80] proposed a recipe to compute an inexact primal-dual search direction so that the equations of the Newton system corresponding to the primal and dual residuals were both satisfied. In the context of CQP, such an approach is no longer possible. Instead, we propose a way to

compute an inexact primal-dual search direction so that the equation corresponding to the primal residual is satisfied exactly, while the one corresponding to the dual residual contains a manageable error which allows us to establish a polynomial bound on the number of outer iterations of our method. Interestingly, the presence of this error on the dual residual equation implies that the primal and dual residuals go to zero at different rates. This is a unique feature of the convergence analysis of our algorithm in that it contrasts with the analysis of other interior-point PDIPF algorithms, where the primal and dual residuals are required to go to zero at the same rate.

The use of inexact search directions in interior-point methods has been extensively studied in the context of cone programming problems (see e.g., [3, 5, 37, 65, 77, 97, 133]). Moreover, the use of iterative linear solvers to compute the primal-dual Newton search directions of interior-point path following algorithms has also been extensively investigated in [3, 6, 12, 37, 65, 97, 99, 102, 111]. For feasibility problems of the form $\{x \in \mathcal{H}_1 : \mathcal{A}x = b, x \in \mathcal{C}\}$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces, $\mathcal{C} \subseteq \mathcal{H}_1$ is a closed convex cone satisfying some mild assumptions, and $\mathcal{A} : \mathcal{H}_1 \to \mathcal{H}_2$ is a continuous linear operator, Renegar [110] has proposed an interior-point method where the Newton system that determines the search directions is approximately solved by performing a uniformly bounded number of iterations of the conjugate gradient (CG) method. To our knowledge, no one has used the ANE system in the context of CQP to obtain either an exact or inexact primal-dual search direction.

This chapter is organized as follows. In Subsection 4.1.1, we give the terminology and notation which will be used throughout this chapter. Section 4.2 describes the outer iteration framework for our algorithm and the complexity results we have obtained for it, along with presenting the ANE as a means to determine the search direction. In Section 4.3, we discuss the use of iterative linear solvers to obtain a suitable approximate solution to the ANE and the construction of an inexact search direction based on this solution. Section 4.4 gives the proofs of the results presented in Sections 4.2 and 4.3. Finally, we present some concluding remarks in Section 4.5.

4.1.1 Terminology and Notation

Throughout this chapter, upper-case Roman letters denote matrices, lower-case Roman letters denote vectors, and lower-case Greek letters denote scalars. We let $\Re^n$, $\Re^n_+$ and $\Re^n_{++}$ denote the set of n- vectors having real, nonnegative real, and positive real components, respectively. Also, we let $\Re^{m\times n}$ denote the set of $m \times n$ matrices with real entries. For a vector $v \in \Re^n$, we let $|v|$ denote the vector whose ith component is $|v_i|$, for every $i = 1, \ldots, n$, and we let $\mathrm{Diag}(v)$ denote the diagonal matrix whose i-th diagonal element is v_i, for every $i = 1, \ldots, n$. In addition, given vectors $u \in \Re^m$ and $v \in \Re^n$, we denote by (u, v) the vector $(u^T, v^T)^T \in \Re^{m+n}$.

Certain matrices bear special notation, namely the matrices X, ΔX, S, D, and $\tilde{D}$. These matrices are the diagonal matrices corresponding to the vectors x, Δx, s, d, and $\tilde{d}$, respectively, as described in the previous paragraph. The symbol 0 will be used to denote a scalar, vector, or matrix of all zeroes; its dimensions should be clear from the context. Also, we denote by e the vector of all 1's, and by I the identity matrix; their dimensions should be clear from the context.

For a symmetric positive definite matrix W, we denote its condition number by $\kappa(W)$, i.e., its maximum eigenvalue divided by its minimum eigenvalue. We will denote sets by upper-case script Roman letters (e.g., B, $\mathcal{N}$). For a finite set B, we denote its cardinality by $|B|$. Given a matrix $A \in \Re^{m\times n}$ and an ordered set $B \subseteq \{1, \ldots, n\}$, we let A_B denote the submatrix whose columns are $\{A_i : i \in B\}$ arranged in the same order as B. Similarly, given a vector $v \in \Re^n$ and an ordered set $B \subseteq \{1, \ldots, n\}$, we let v_B denote the subvector consisting of the elements $\{v_i : i \in B\}$ arranged in the same order as B.

We will use several different norms throughout this chapter. For a vector $z \in \Re^n$, $\|z\| = \sqrt{z^T z}$ is the Euclidian norm, $\|z\|_1 = \sum_{i=1}^n |z_i|$ is the "1-norm", and $\|z\|_\infty = \max_{i=1,\ldots,n} |z_i|$ is the "infinity norm". For a matrix $V \in \Re^{m\times n}$, $\|V\|$ denotes the operator norm associated with the Euclidian norm: $\|V\| = \max_{z:\|z\|=1} \|Vz\|$. Finally, $\|V\|_F$ denotes the Frobenius norm: $\|V\|_F = (\sum_{i=1}^m \sum_{j=1}^n V_{ij}^2)^{1/2}$.

4.2 Outer Iteration Framework

In this section, we introduce our PDIPF algorithm based on a class of inexact search directions and discuss its iteration complexity. This section is divided into two subsections. In Subsection 4.2.1, we discuss an exact PDIPF algorithm, which will serve as the basis for the inexact PDIPF algorithm given in Subsection 4.2.2, and we give its iteration complexity result. We also present an approach based on the ANE to determine the Newton search direction for the exact algorithm. To motivate the class of inexact search directions used by our inexact PDIPF algorithm, we describe in Subsection 4.2.2 a framework for computing an inexact search direction based on an approximate solution to the ANE. We then introduce the class of inexact search directions, state a PDIPF algorithm based on it, and give its iteration complexity result.

4.2.1 An Exact PDIPF Algorithm and the ANE

Consider the following primal-dual pair of CQP problems:

$$\min_x \quad \left\{\frac{1}{2}x^T V E^2 V^T x + c^T x : \quad Ax = b,\ x \geq 0\right\}, \tag{135}$$

$$\max_{(\hat{x},s,y)} \quad \left\{-\frac{1}{2}\hat{x}^T V E^2 V^T \hat{x} + b^T y : \quad A^T y + s - V E^2 V^T \hat{x} = c,\ s \geq 0\right\}, \tag{136}$$

where the data are $V \in \Re^{n\times l}$, $E \in \mathrm{Diag}(\Re^l_{++})$, $A \in \Re^{m\times n}$, $b \in \Re^m$ and $c \in \Re^n$, and the decision variables are $x \in \Re^n$ and $(\hat{x}, s, y) \in \Re^n \times \Re^n \times \Re^m$. We observe that the Hessian matrix Q is already given in factored form $Q = VE^2V^T$.

It is well-known that if x^* is an optimal solution for (135) and $(\hat{x}^*, s^*, y^*)$ is an optimal solution for (136), then (x^*, s^*, y^*) is also an optimal solution for (136). Now, let $\mathcal{S}$ denote the set of all vectors $w := (x, s, y, z) \in \Re^{2n+m+l}$ satisfying

$$Ax = b, \quad x \geq 0, \tag{137}$$

$$A^T y + s + Vz = c, \quad s \geq 0, \tag{138}$$

$$Xs = 0, \tag{139}$$

$$EV^T x + E^{-1} z = 0. \tag{140}$$

It is clear that $w \in \mathcal{S}$ if and only if x is optimal for (135), (x, s, y) is optimal for (136),

and $z = -E^2V^Tx$. (Throughout this chapter, the symbol w will always denote the quadruple (x, s, y, z), where the vectors lie in the appropriate dimensions; similarly, $\Delta w = (\Delta x, \Delta s, \Delta y, \Delta z)$, $w^k = (x^k, s^k, y^k, z^k)$, $\bar{w} = (\bar{x}, \bar{s}, \bar{y}, \bar{z})$, etc.)

We observe that the presentation of the PDIPF algorithm based on exact Newton search directions in this subsection differs from the classical way of presenting it in that we introduce an additional variable z as above. Clearly, it is easy to see that the variable z is completely redundant and can be eliminated, thereby reducing the method described below to the usual way of presenting it. The main reason for introducing the variable z is due to the development of the ANE presented at the end of this subsection.

We will make the following two assumptions throughout this chapter:

Assumption 1 *A has full row rank.*

Assumption 2 *The set $\mathcal{S}$ is nonempty.*

For a point $w \in \Re^{2n}_{++} \times \Re^{m+l}$, let us define

$$\mu := \mu(w) = x^Ts/n, \tag{141}$$

$$r_p := r_p(w) = Ax - b, \tag{142}$$

$$r_d := r_d(w) = A^Ty + s + Vz - c, \tag{143}$$

$$r_V := r_V(w) = EV^Tx + E^{-1}z, \tag{144}$$

$$r := r(w) = (r_p(w), r_d(w), r_V(w)). \tag{145}$$

Moreover, given $\gamma \in (0, 1)$ and an initial point $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$, we define the following neighborhood of the central path:

$$\mathcal{N}_{w^0}(\gamma) := \left\{ w \in \Re^{2n}_{++} \times \Re^{m+l} : Xs \geq (1-\gamma)\mu e,\ r = \eta r^0,\ 0 \leq \eta \leq \min\left[1, \frac{\mu}{\mu_0}\right] \right\} \tag{146}$$

where $r := r(w)$, $r^0 := r(w^0)$, $\mu := \mu(w)$, and $\mu_0 := \mu(w^0)$.

We are now ready to state the PDIPF algorithm based on exact Newton search directions.

Exact PDIPF Algorithm

1. **Start:** Let $\epsilon > 0$ and $0 < \underline{\sigma} \leq \overline{\sigma} < 1$ be given. Let $\gamma \in (0,1)$ and $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ be such that $w^0 \in \mathcal{N}_{w^0}(\gamma)$. Set $k = 0$.

2. **While** $\mu_k := \mu(w^k) > \epsilon$ **do**

   (a) Let $w := w^k$ and $\mu := \mu_k$; choose $\sigma := \sigma_k \in [\underline{\sigma}, \overline{\sigma}]$.

   (b) Let $\Delta w = (\Delta x, \Delta s, \Delta y, \Delta z)$ denote the solution of the linear system

$$A\Delta x = -r_p, \tag{147}$$

$$A^T\Delta y + \Delta s + V\Delta z = -r_d, \tag{148}$$

$$X\Delta s + S\Delta x = -Xs + \sigma\mu e, \tag{149}$$

$$EV^T\Delta x + E^{-1}\Delta z = -r_V. \tag{150}$$

   (c) Let $\tilde{\alpha} = \text{argmax}\,\{\alpha \in [0,1] : w + \alpha'\Delta w \in \mathcal{N}_{w^0}(\gamma),\ \forall \alpha' \in [0,\alpha]\}$.

   (d) Let $\bar{\alpha} = \text{argmin}\left\{(x + \alpha\Delta x)^T(s + \alpha\Delta s) : \alpha \in [0, \tilde{\alpha}]\right\}$.

   (e) Let $w^{k+1} = w + \bar{\alpha}\Delta w$, and set $k \leftarrow k + 1$.

   End (while)

A proof of the following result, under slightly different assumptions, can be found in [132].

Theorem 4.2.1 *Assume that the constants γ, $\underline{\sigma}$, and $\overline{\sigma}$ are such that*

$$\max\left\{\gamma^{-1}, (1-\gamma)^{-1}, \underline{\sigma}^{-1}, (1-\overline{\sigma})^{-1}\right\} = \mathcal{O}(1),$$

and that the initial point $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfies $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Then, the Exact PDIPF Algorithm finds an iterate $w^k \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfying $\mu_k \leq \epsilon\mu_0$ and $\|r^k\| \leq \epsilon\|r^0\|$ within $\mathcal{O}(n^2\log(1/\epsilon))$ iterations.

A few approaches have been suggested in the literature for computing the Newton search direction (147)-(150). Instead of using one of them, we will discuss below a new approach, referred to in this chapter as the ANE approach, that we believe to be suitable not only for direct solvers but especially for iterative linear solvers as we will see in Section 4.3.

Let us begin by defining the following matrices:

$$D := X^{1/2}S^{-1/2}, \tag{151}$$

$$\tilde{D} := \begin{pmatrix} D & 0 \\ 0 & E^{-1} \end{pmatrix} \in \Re^{(n+l)\times(n+l)}, \tag{152}$$

$$\tilde{A} := \begin{pmatrix} A & 0 \\ V^T & I \end{pmatrix} \in \Re^{(m+l)\times(n+l)}. \tag{153}$$

Suppose that we first solve the following system of equations for $(\Delta y, \Delta z)$:

$$\tilde{A}\tilde{D}^2\tilde{A}^T \begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} = \tilde{A} \begin{pmatrix} x - \sigma\mu S^{-1}e - D^2 r_d \\ 0 \end{pmatrix} + \begin{pmatrix} -r_p \\ -E^{-1}r_V \end{pmatrix} =: h. \tag{154}$$

This system is what we refer to as the ANE. Next, we obtain Δs and Δx according to:

$$\Delta s = -r_d - A^T\Delta y - V\Delta z, \tag{155}$$

$$\Delta x = -D^2\Delta s - x + \sigma\mu S^{-1}e. \tag{156}$$

Clearly, the search direction $\Delta w = (\Delta x, \Delta s, \Delta y, \Delta z)$ computed as above satisfies (148) and (149) in view of (155) and (156). Moreover, it also satisfies (147) and (150) due to the fact that by (152), (153), (154), (155) and (156), we have that

$$\begin{aligned}
\tilde{A}\begin{pmatrix} \Delta x \\ E^{-2}\Delta z \end{pmatrix} &= \tilde{A}\begin{pmatrix} -D^2\Delta s - x + \sigma\mu S^{-1}e \\ E^{-2}\Delta z \end{pmatrix} \\
&= \tilde{A}\begin{pmatrix} D^2 r_d + D^2A^T\Delta y + D^2V\Delta z - x + \sigma\mu S^{-1}e \\ E^{-2}\Delta z \end{pmatrix} \\
&= \tilde{A}\begin{pmatrix} D^2A^T\Delta y + D^2V\Delta z \\ E^{-2}\Delta z \end{pmatrix} + \tilde{A}\begin{pmatrix} D^2 r_d - x + \sigma\mu S^{-1}e \\ 0 \end{pmatrix} \\
&= \tilde{A}\tilde{D}^2\tilde{A}^T\begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} + \tilde{A}\begin{pmatrix} D^2 r_d - x + \sigma\mu S^{-1}e \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -r_p \\ -E^{-1}r_V \end{pmatrix}.
\end{aligned} \tag{157}$$

Theorem 4.2.1 assumes that Δw is the exact solution of (154), which is usually obtained by computing the Cholesky factorization of the coefficient matrix of the ANE. In this

chapter, we will consider a variant of the Exact PDIPF Algorithm whose search directions are approximate solutions of (154) and ways of determining these inexact search directions by means of a suitable preconditioned iterative linear solver.

4.2.2 An Inexact PDIPF algorithm for CQP

In this subsection, we describe a PDIPF algorithm based on a family of search directions that are approximate solutions to (147)–(150) and discuss its iteration complexity properties.

Clearly, an approximate solution to the ANE can only yield an approximate solution to (147)–(150). In order to motivate the class of inexact search directions used by the PDIPF algorithm presented in this subsection, we present a framework for obtaining approximate solutions to (147)–(150) based on an approximate solution to the ANE.

Suppose that the ANE is solved only inexactly, i.e., that the vector $(\Delta y, \Delta z)$ satisfies

$$\tilde{A}\tilde{D}^2\tilde{A}^T \begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} = h + f \tag{158}$$

for some error vector f. If Δs and Δx were computed by (155) and (156), respectively, then it is clear that the search direction Δw would satisfy (148) and (149). However, (147) and (150) would not be satisfied, since by an argument similar to (157), we would have that

$$\begin{aligned} \tilde{A} \begin{pmatrix} \Delta x \\ E^{-2}\Delta z \end{pmatrix} &= \ldots = \tilde{A}\tilde{D}^2\tilde{A}^T \begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} + \tilde{A} \begin{pmatrix} D^2 r_d - x + \sigma\mu S^{-1} e \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -r_p \\ -E^{-1} r_V \end{pmatrix} + f. \end{aligned}$$

Instead, suppose we use (155) to determine Δs as before, but now we determine Δx as

$$\Delta x = -D^2\Delta s - x + \sigma\mu S^{-1} e - S^{-1} p, \tag{159}$$

where the correction vector $p \in \Re^n$ will be required to satisfy some conditions which we will now describe.

To motivate the conditions on p, we note that (155), (158), and (159) imply that

$$\begin{aligned}
&\tilde{A}\begin{pmatrix} \Delta x \\ E^{-2}\Delta z \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} = \tilde{A}\begin{pmatrix} -D^2\Delta s - x + \sigma\mu S^{-1}e - S^{-1}p \\ E^{-2}\Delta z \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\
&= \tilde{A}\begin{pmatrix} D^2 r_d + D^2 A^T \Delta y + D^2 V \Delta z - x + \sigma\mu S^{-1}e - S^{-1}p \\ E^{-2}\Delta z \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\
&= \tilde{A}\tilde{D}^2\begin{pmatrix} A^T\Delta y + V\Delta z \\ \Delta z \end{pmatrix} + \tilde{A}\begin{pmatrix} D^2 r_d - x + \sigma\mu S^{-1}e \\ 0 \end{pmatrix} - \tilde{A}\begin{pmatrix} S^{-1}p \\ 0 \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\
&= \tilde{A}\tilde{D}^2\tilde{A}^T\begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} + \tilde{A}\begin{pmatrix} D^2 r_d - x + \sigma\mu S^{-1}e \\ 0 \end{pmatrix} - \tilde{A}\begin{pmatrix} S^{-1}p \\ 0 \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\
&= f - \tilde{A}\begin{pmatrix} S^{-1}p \\ 0 \end{pmatrix}.
\end{aligned} \tag{160}$$

Based on the above equation, one is naturally tempted to choose p so that the right hand side of (160) is zero, and consequently (147) and (150) are satisfied exactly. However, the existence of such p cannot be guaranteed and, even if it exists, its magnitude might not be sufficiently small to yield a search direction which is suitable for the development of a polynomially convergent algorithm. Instead, we consider an alternative approach where p is chosen so that the first component of (160) is zero and the second component is small. More specifically, by partitioning $f = (f_1, f_2) \in \Re^m \times \Re^l$, we choose $p \in \Re^n$ such that

$$AS^{-1}p = f_1. \tag{161}$$

It is clear that p is not uniquely defined. Note that (153) implies that (161) is equivalent to

$$f \;=\; \tilde{A}\begin{pmatrix} S^{-1}p \\ E^{-1}q \end{pmatrix}, \tag{162}$$

where $q := E(f_2 - V^T S^{-1}p)$. Then, using (153), (160), and (162), we conclude that

$$\begin{aligned}
\tilde{A}\begin{pmatrix} \Delta x \\ E^{-2}\Delta z \end{pmatrix} + \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \;&=\; f - \tilde{A}\begin{pmatrix} S^{-1}p \\ E^{-1}q \end{pmatrix} + \tilde{A}\begin{pmatrix} 0 \\ E^{-1}q \end{pmatrix} \\
&=\; \tilde{A}\begin{pmatrix} 0 \\ E^{-1}q \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ E^{-1}q \end{pmatrix},
\end{aligned} \tag{163}$$

from which we see that the first component of (160) is set to 0 and the second component is exactly $E^{-1}q$.

In view of (155), (159), and (163), the above construction yields a search direction Δw satisfying the following modified Newton system of equations:

$$A\Delta x = -r_p, \tag{164}$$

$$A^T\Delta y + \Delta s + V\Delta z = -r_d, \tag{165}$$

$$X\Delta s + S\Delta x = -Xs + \sigma\mu e - p, \tag{166}$$

$$EV^T\Delta x + E^{-1}\Delta z = -r_V + q. \tag{167}$$

As far as the outer iteration complexity analysis of our algorithm is concerned, all we require of our inexact search directions is that they satisfy (164)–(167) and that p and q be relatively small in the following sense:

Definition 2 *Given a point $w \in \Re^{2n}_{++} \times \Re^{m+l}$ and positive scalars τ_p and τ_q, an inexact direction Δw is referred to as a* (τ_p, τ_q)-search direction *if it satisfies (164)–(167) for some p and q satisfying $\|p\|_\infty \leq \tau_p\mu$ and $\|q\| \leq \tau_q\sqrt{\mu}$, where μ is given by (141).*

We next define a generalized central path neighborhood which is used by our inexact PDIPF algorithm. Given a starting point $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ and parameters $\eta \geq 0$, $\gamma \in [0,1]$, and $\theta > 0$, define the following set:

$$\mathcal{N}_{w^0}(\eta,\gamma,\theta) = \left\{ w \in \Re^{2n}_{++} \times \Re^{m+l} : \begin{array}{cc} Xs \geq (1-\gamma)\mu e, & (r_p, r_d) = \eta(r_p^0, r_d^0), \\ \|r_V - \eta r_V^0\| \leq \theta\sqrt{\mu}, & \eta \leq \mu/\mu_0 \end{array} \right\}, \tag{168}$$

where $\mu = \mu(w)$, $\mu_0 = \mu(w^0)$, $r = r(w)$ and $r^0 = r(w^0)$. The generalized central path neighborhood is then given by

$$\mathcal{N}_{w^0}(\gamma,\theta) = \bigcup_{\eta\in[0,1]} \mathcal{N}_{w^0}(\eta,\gamma,\theta). \tag{169}$$

We observe that the neighborhood given by (169) agrees with the neighborhood given by (146) when $\theta = 0$.

We are now ready to state our inexact PDIPF algorithm.

Inexact PDIPF Algorithm:

1. **Start:** Let $\epsilon > 0$ and $0 < \underline{\sigma} \leq \overline{\sigma} < 4/5$ be given. Choose $\gamma \in (0,1)$, $\theta > 0$ and $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ such that $w^0 \in \mathcal{N}_{w^0}(\gamma, \theta)$. Set $k = 0$.

2. **While** $\mu_k := \mu(w^k) > \epsilon$ **do**

   (a) Let $w := w^k$ and $\mu := \mu_k$; choose $\sigma \in [\underline{\sigma}, \overline{\sigma}]$.

   (b) Set

$$\tau_p = \gamma\sigma/4 \quad \text{and} \tag{170}$$

$$\tau_q = \left[\sqrt{1 + (1 - 0.5\gamma)\,\sigma} - 1\right]\theta. \tag{171}$$

   (c) Set $r_p = Ax - b$, $r_d = A^T y + s + Vz - c$, $r_V = EV^T x + E^{-1}z$, and $\eta = \|r_p\|/\|r_p^0\|$.

   (d) Compute a (τ_p, τ_q)-search direction Δw.

   (e) Compute $\tilde{\alpha} := \text{argmax}\{\alpha \in [0,1] : w + \alpha'\Delta w \in \mathcal{N}_{w^0}(\gamma, \theta),\ \forall \alpha' \in [0, \alpha]\}$.

   (f) Compute $\bar{\alpha} := \text{argmin}\{(x + \alpha\Delta x)^T(s + \alpha\Delta s) : \alpha \in [0, \tilde{\alpha}]\}$.

   (g) Let $w^{k+1} = w + \bar{\alpha}\Delta w$, and set $k \leftarrow k + 1$.

   End (while)

The following result gives a bound on the number of iterations needed by the Inexact PDIPF Algorithm to obtain an ϵ-solution to the KKT conditions (137)–(140). Its proof will be given in Subsection 4.4.2.

Theorem 4.2.1 *Assume that the constants γ, $\underline{\sigma}$, $\overline{\sigma}$ and θ are such that*

$$\max\left\{\gamma^{-1}\,,\,(1-\gamma)^{-1}\,,\,\underline{\sigma}^{-1}\,,\,\left(1 - \frac{5}{4}\overline{\sigma}\right)^{-1}\right\} = \mathcal{O}(1),\ \ \theta = \mathcal{O}(\sqrt{n}), \tag{172}$$

and that the initial point $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfies $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Then, the Inexact PDIPF Algorithm generates an iterate $w^k \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfying $\mu_k \leq \epsilon\mu_0$, $\|(r_p^k, r_d^k)\| \leq \epsilon\|(r_p^0, r_d^0)\|$, and $\|r_V^k\| \leq \epsilon\|r_V^0\| + \epsilon^{1/2}\theta\mu_0^{1/2}$ within $\mathcal{O}\left(n^2\log(1/\epsilon)\right)$ iterations.

4.3 Determining an Inexact Search Direction Via an Iterative Solver

The results in Subsection 4.2.2 assume we can obtain a (τ_p, τ_q)-search direction Δw, where τ_p and τ_q are given by (170) and (171), respectively. In this section, we will describe a way to obtain a (τ_p, τ_q)-search direction Δw using a uniformly bounded number of iterations of a suitable preconditioned iterative linear solver applied to the ANE. It turns out that the construction of this Δw is based on the recipe given at the beginning of Subsection 4.2.2, together with a specific choice of the perturbation vector p.

This section is divided into two subsections. In Subsection 4.3.1, we introduce the MWB preconditioner which will be used to precondition the ANE. In addition, we also introduce a family of iterative linear solvers used to solve the preconditioned ANE. Subsection 4.3.2 gives a specific approach for constructing a pair (p, q) satisfying (162), and an approximate solution to the ANE so that the recipe described at the beginning of Subsection 4.2.2 yields a (τ_p, τ_q)-search direction Δw. It also provides a uniform bound on the number of iterations that any member of the family of iterative linear solvers needs to perform to obtain such a direction Δw when applied to the preconditioned ANE.

4.3.1 MWB Preconditioner and a Family of Solvers

In this subsection we introduce the MWB preconditioner, and we discuss its use as a preconditioner in solving the ANE via a family of iterative linear solvers. Since the condition number of the ANE matrix $\tilde{A}\tilde{D}^2\tilde{A}^T$ may "blow up" for points w near an optimal solution, the direct application of a generic iterative linear solver for solving the ANE without first preconditioning it is generally not effective. We discuss a natural remedy to this problem which consists of using a preconditioner $\tilde{T}$, namely the MWB preconditioner, such that $\kappa(\tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T)$ remains uniformly bounded regardless of the iterate w. Finally, we analyze the complexity of the resulting approach to obtain a suitable approximate solution to the ANE.

We start by describing the MWB preconditioner. Its construction essentially consists of building a basis B of $\tilde{A}$ which gives higher priority to the columns of $\tilde{A}$ corresponding

to larger diagonal elements of $\tilde{D}$. More specifically, the MWB preconditioner is determined by the following algorithm:

Maximum Weight Basis Algorithm

Start: Given $\tilde{d} \in \Re_{++}^{(n+l)}$, and $\tilde{A} \in \Re^{(m+l)\times(n+l)}$ such that $\text{rank}(\tilde{A}) = m + l$,

1. Order the elements of $\tilde{d}$ so that $\tilde{d}_1 \geq \ldots \geq \tilde{d}_{n+l}$; order the columns of $\tilde{A}$ accordingly.

2. Let $B = \emptyset$, $j = 1$.

3. **While** $|B| < m + l$ **do**

   (a) If $\tilde{A}_j$ is linearly independent of $\{\tilde{A}_i : i \in B\}$, set $B \leftarrow B \cup \{j\}$.

   (b) $j \leftarrow j + 1$.

4. Return to the original ordering of $\tilde{A}$ and $\tilde{d}$; determine the set B according to this ordering and set $\mathcal{N} := \{1, \ldots, n + l\} \backslash B$.

5. Set $B := \tilde{A}_B$ and $\tilde{D}_B := \text{Diag}(\tilde{d}_B)$.

6. Let $\tilde{T} = \tilde{T}(\tilde{A}, \tilde{d}) := \tilde{D}_B^{-1} B^{-1}$.

end

Note that the above algorithm can be applied to the matrix $\tilde{A}$ defined in (153) since this matrix has full row rank due to Assumption 1. The MWB preconditioner was originally proposed by Vaidya [124] and Resende and Veiga [111] in the context of the minimum cost network flow problem. In this case, $\tilde{A} = A$ is the node-arc incidence matrix of a connected digraph (with one row deleted to ensure that $\tilde{A}$ has full row rank), the entries of $\tilde{d}$ are weights on the edges of the graph, and the set B generated by the above algorithm defines a maximum spanning tree on the digraph. Oliveira and Sorensen [99] later proposed the use of this preconditioner for general matrices $\tilde{A}$. Boman et. al. [14] have proposed variants of the MWB preconditioner for diagonally dominant matrices, using the fact that they can be represented as $D_1 + AD_2A^T$, where D_1 and D_2 are nonnegative diagonal and positive diagonal matrices, respectively, and A is a node-arc incidence matrix.

For the purpose of stating the next result, we now introduce some notation. Let us define

$$\varphi_{\tilde{A}} \;:=\; \max\{\|B^{-1}\tilde{A}\|_F : B \text{ is a basis of } \tilde{A}\}. \tag{173}$$

The constant $\varphi_{\tilde{A}}$ is related to the well-known condition number $\bar{\chi}_{\tilde{A}}$ (see [126]), defined as

$$\bar{\chi}_{\tilde{A}} \;:=\; \sup\{\|\tilde{A}^T(\tilde{A}\tilde{E}\tilde{A}^T)^{-1}\tilde{A}\tilde{E}\| : \tilde{E} \in \mathrm{Diag}(\Re_{++}^{(n+l)})\}.$$

Specifically, $\varphi_{\tilde{A}} \leq (n+l)^{1/2}\bar{\chi}_{\tilde{A}}$, in view of the facts that $\|C\|_F \leq (n+l)^{1/2}\,\|C\|$ for any matrix $C \in \Re^{(m+l)\times(n+l)}$ and, as shown in [123] and [126],

$$\bar{\chi}_{\tilde{A}} \;=\; \max\{\|B^{-1}\tilde{A}\| : B \text{ is a basis of } \tilde{A}\}.$$

The following result, which establishes the theoretical properties of the MWB preconditioner, follows as a consequence of Lemmas 2.1 and 2.2 of [81].

Proposition 4.3.1 *Let $\tilde{T} = \tilde{T}(\tilde{A}, \tilde{d})$ be the preconditioner determined according to the Maximum Weight Basis Algorithm, and define $W := \tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T$. Then, $\|\tilde{T}\tilde{A}\tilde{D}\| \leq \varphi_{\tilde{A}}$ and $\kappa(W) \leq \varphi_{\tilde{A}}^2$.*

Note that the bound $\varphi_{\tilde{A}}^2$ on $\kappa(W)$ is independent of the diagonal matrix $\tilde{D}$ and depends only on $\tilde{A}$. This will allow us to obtain a uniform bound on the number of iterations needed by any member of the family of iterative linear solvers described below to obtain a suitable approximate solution of (154). This topic is the subject of the remainder of this subsection.

Instead of dealing directly with (154), we consider the application of an iterative linear solver to the preconditioned ANE:

$$Wu = v, \tag{174}$$

where

$$W := \tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T, \qquad v := \tilde{T}h. \tag{175}$$

For the purpose of our analysis below, the only thing we will assume regarding the iterative linear solver when applied to (174) is that it generates a sequence of iterates $\{u^j\}$ such that

$$\|v - Wu^j\| \;\leq\; c(\kappa)\left[1 - \frac{1}{\psi(\kappa)}\right]^j \|v - Wu^0\|, \;\; \forall\; j = 0, 1, 2, \ldots, \tag{176}$$

Table 3: Values of $c(\kappa)$ and $\psi(\kappa)$ for well-known Iterative Solvers

Solver	$c(\kappa)$	$\psi(\kappa)$
SD	$\sqrt{\kappa}$	$(\kappa+1)/2$
CG	$2\sqrt{\kappa}$	$(\sqrt{\kappa}+1)/2$

where c and ψ are positive, nondecreasing functions of $\kappa \equiv \kappa(W)$.

Examples of solvers which satisfy (176) include the steepest descent (SD) and CG methods, with the following values for $c(\kappa)$ and $\psi(\kappa)$:

The justification for the table above follows from Section 7.6 and Exercise 10 of Section 8.8 of [71].

The following result gives an upper bound on the number of iterations required by any iterative linear solver satisfying (176) needs to perform to obtain a *ξ-approximate solution of (174)*, i.e., an iterate u^j such that $\|v - Wu^j\| \leq \xi\sqrt{\mu}$ for some constant $\xi > 0$:

Proposition 4.3.2 *Let u^0 be an arbitrary starting point. Then, a generic iterative linear solver with a convergence rate given by (176) generates an iterate u^j satisfying $\|v - Wu^j\| \leq \xi\sqrt{\mu}$ in*

$$\mathcal{O}\left(\psi(\kappa)\log\left(\frac{c(\kappa)\|v - Wu^0\|}{\xi\sqrt{\mu}}\right)\right) \tag{177}$$

iterations, where $\kappa \equiv \kappa(W)$.

Proof. Let j be any iteration such that $\|v - Wu^j\| > \xi\sqrt{\mu}$. We use relation (176) and the fact that $1 + \omega \leq e^{\omega}$ for all $\omega \in \Re$ to observe that

$$\xi\sqrt{\mu} \;<\; \|v - Wu^j\| \;\leq\; c(\kappa)\left[1 - \frac{1}{\psi(\kappa)}\right]^j \|v - Wu^0\| \;\leq\; c(\kappa)\exp\left\{\frac{-j}{\psi(\kappa)}\right\}\|v - Wu^0\|.$$

Rearranging the first and last terms of the inequality, it follows that

$$j \;<\; \psi(\kappa)\log\left(\frac{c(\kappa)\|v - Wu^0\|}{\xi\sqrt{\mu}}\right),$$

and the result is proven. ∎

From Proposition 4.3.2, it is clear that different choices of u^0 and ξ lead to different bounds on the number of iterations performed by the iterative linear solver. In Subsection

4.3.2, we will describe a suitable way of selecting u^0 and ξ so that (i) the bound (177) is independent of the iterate w and (ii) the approximate solution $\tilde{T}^T u^j$ of the ANE, together with a suitable pair (p, q), yields a (τ_p, τ_q)-search direction Δw through the recipe described in Subsection 4.2.2.

4.3.2 Computation of the Inexact Search Direction Δw

In this subsection, we use the results of Subsections 4.2.2 and 4.3.1 to build a (τ_p, τ_q)-search direction Δw, where τ_p and τ_q are given by (170) and (171), respectively. In addition, we describe a way of choosing u^0 and ξ which ensures that the number of iterations of an iterative linear solver satisfying (176) applied to the preconditioned ANE is uniformly bounded by a constant depending on n and $\varphi_{\tilde{A}}$.

Suppose that we solve (174) inexactly according to Subsection 4.3.1. Then our final solution u^j satisfies $Wu^j - v = \tilde{f}$ for some vector $\tilde{f}$. Letting

$$\begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} = \tilde{T}^T u^j, \tag{178}$$

we easily see from (175) that (158) is satisfied with $f := \tilde{T}^{-1}\tilde{f}$. We can then apply the recipe of Subsection 4.2.2 to this approximate solution, using the pair (p, q) which we will now describe.

First, note that (162) with f as defined above is equivalent to the system

$$\tilde{f} \;=\; \tilde{T}\tilde{A}\begin{pmatrix} S^{-1}p \\ E^{-1}q \end{pmatrix} \;=\; \tilde{T}\tilde{A}\tilde{D}\begin{pmatrix} (XS)^{-1/2} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix}. \tag{179}$$

Now, let $B = (B_1, \ldots, B_{m+l})$ be the ordered set of basic indices computed by the MWB Algorithm applied to the pair $(\tilde{A}, \tilde{d})$ and note that, by step 6 of this algorithm, the B_i-th column of $\tilde{T}\tilde{A}\tilde{D}$ is the ith unit vector for every $i = 1, \ldots, m + l$. Then, the vector $t \in \Re^{n+l}$ defined as $t_{B_i} = \tilde{f}_i$ for $i = 1, \ldots, m + l$ and $t_j = 0$ for every $j \notin \{B_1, \ldots, B_{m+l}\}$ clearly satisfies

$$\tilde{f} \;=\; \tilde{T}\tilde{A}\tilde{D}\, t. \tag{180}$$

We then obtain a pair $(p, q) \in \Re^n \times \Re^l$ satisfying (162) by defining

$$\begin{pmatrix} p \\ q \end{pmatrix} := \begin{pmatrix} (XS)^{1/2} & 0 \\ 0 & I \end{pmatrix} t. \tag{181}$$

It is clear from (181) and the fact that $\|t\| = \|\tilde{f}\|$ that

$$\|p\| \;\le\; \|XS\|^{1/2}\|\tilde{f}\|, \qquad \|q\| \;\le\; \|\tilde{f}\|. \tag{182}$$

As an immediate consequence of this relation, we obtain the following result.

Lemma 4.3.3 *Suppose that $w \in \Re^{2n}_{++} \times \Re^{m+l}$ and positive scalars τ_p and τ_q are given. Assume that u^j is a ξ-approximate solution of (174), or equivalently $\tilde{f} \le \xi\sqrt{\mu}$, where $\xi := \min\{n^{-1/2}\tau_p, \tau_q\}$. Let Δw be determined according to the recipe given in Subsection 4.2.2 using the approximate solution (178) and the pair (p, q) given by (181). Then Δw is a (τ_p, τ_q)-search direction.*

Proof. It is clear from the previous discussion that Δw and the pair (p, q) satisfy (164)–(167). Next, relation (182) and the facts that $\xi \le n^{-1/2}\tau_p$ and $\|XS\|^{1/2} \le \sqrt{n\mu}$ imply that

$$\|p\|_\infty \;\le\; \|p\| \;\le\; \|XS\|^{1/2}\|\tilde{f}\| \;\le\; \sqrt{n\mu}\,\xi\sqrt{\mu} \;\le\; \tau_p\mu.$$

Similarly, (182) and the fact that $\xi \le \tau_q$ imply that $\|q\| \le \tau_q\sqrt{\mu}$. Thus, Δw is a (τ_p, τ_q)-search direction as desired. ∎

Lemma (4.3.3) implies that, to construct a (τ_p, τ_q)-search direction Δw as in step 2(d) of the Inexact PDIPF Algorithm, it suffices to find a ξ-approximate solution to (174), where

$$\xi \;:=\; \min\left\{\frac{\gamma\sigma}{4\sqrt{n}}\,,\; \left[\sqrt{1+\left(1-\frac{\gamma}{2}\right)\sigma}-1\right]\theta\right\}. \tag{183}$$

We next describe a suitable way of selecting u^0 so that the number of iterations required by an iterative linear solver satisfying (176) to find a ξ-approximate solution of (174) can be uniformly bounded by a universal constant depending only on the quantities n and $\varphi_{\tilde{A}}$. First, compute a point $\bar{w} = (\bar{x}, \bar{s}, \bar{y}, \bar{z})$ such that

$$\tilde{A}\begin{pmatrix} \bar{x} \\ E^{-2}\bar{z} \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}, \qquad A^T\bar{y} + \bar{s} + V\bar{z} = c. \tag{184}$$

Note that vectors $\bar{x}$ and $\bar{z}$ satisfying the first equation in (184) can be easily computed once a basis of $\tilde{A}$ is available (e.g., the one computed by the Maximum Weight Basis Algorithm in the first outer iteration of the Inexact PDIPF Algorithm). Once $\bar{y}$ is arbitrarily chosen, a vector $\bar{s}$ satisfying the second equation of (184) is immediately available. We then define

$$u^0 \;=\; -\eta\,\tilde{T}^{-T}\begin{pmatrix} y^0-\bar{y} \\ z^0-\bar{z} \end{pmatrix}. \tag{185}$$

The following lemma gives a bound on the size of the initial residual $\|Wu^0 - v\|$. Its proof will be given in Subsection 4.4.1.

Lemma 4.3.4 *Assume that $\tilde{T} = \tilde{T}(\tilde{A}, \tilde{d})$ is given and that $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ and $\bar{w}$ are such that $(x^0, s^0) \geq |(\bar{x}, \bar{s})|$ and $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Further, assume that $w \in \mathcal{N}_{w^0}(\gamma, \theta)$ for some $\gamma \in [0, 1]$ and $\theta > 0$, and that W, v and u^0 are given by (175) and (185), respectively. Then, the initial residual in (176) satisfies $\|v - Wu^0\| \leq \Psi\sqrt{\mu}$, where*

$$\Psi \;:=\; \left[\frac{7n + \theta^2/2}{\sqrt{1-\gamma}} + \theta\right]\varphi_{\tilde{A}}. \tag{186}$$

As an immediate consequence of Proposition 4.3.2 and Lemmas 4.3.3 and 4.3.4, we can bound the number of inner iterations required by an iterative linear solver satisfying (176) to yield a (τ_p, τ_q)-search direction Δw.

Theorem 4.3.5 *Assume that ξ is defined in (183), where σ, γ, θ are such that*

$$\max\{\sigma^{-1}, \gamma^{-1}, (1-\gamma)^{-1}, \theta, \theta^{-1}\}$$

is bounded by a polynomial of n. Assume also that $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ and $\bar{w}$ are such that $(x^0, s^0) \geq |(\bar{x}, \bar{s})|$ and $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Then, a generic iterative linear solver with a convergence rate given by (176) generates a ξ-approximate solution, which leads to a (τ_p, τ_q)-search direction Δw, in

$$\mathcal{O}\left(\psi(\varphi^2_{\tilde{A}})\log\left(c(\varphi^2_{\tilde{A}})n\varphi_{\tilde{A}}\right)\right) \tag{187}$$

iterations. As a consequence, the SD and CG methods generate this approximate solution u^j *in* $\mathcal{O}(\varphi_{\tilde{A}}^2 \log(n\varphi_{\tilde{A}}))$ *and* $\mathcal{O}(\varphi_{\tilde{A}} \log(n\varphi_{\tilde{A}}))$ *iterations, respectively.*

Proof. The proof of the first part of Theorem 4.3.5 immediately follows from Propositions 4.3.1 and 4.3.2 and Lemmas 4.3.3 and 4.3.4. The proof of the second part of Theorem 4.3.5 follows immediately from Table 3 and Proposition 4.3.1. ∎

Using the results of Sections 4.2 and 4.3, we see that the number of "inner" iterations of an iterative linear solver satisfying (176) is uniformly bounded by a constant depending on n and $\varphi_{\tilde{A}}$, while the number of "outer" iterations in the Inexact PDIPF Algorithm is polynomially bounded by a constant depending on n and $\log\epsilon^{-1}$.

4.4 Technical Results

This section is devoted to the proofs of Lemma 4.3.4 and Theorem 4.2.1. Subsection 4.4.1 presents the proof of Lemma 4.3.4, and Subsection 4.4.2 presents the proof of Theorem 4.2.1.

4.4.1 Proof of Lemma 4.3.4

In this subsection, we will provide the proof of Lemma 4.3.4. We begin by establishing three technical lemmas.

Lemma 4.4.1 *Suppose that* $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$, $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$ *for some* $\eta \in [0,1]$, $\gamma \in [0,1]$ *and* $\theta > 0$, *and* $w^* \in \mathcal{S}$. *Then*

$$(x - \eta x^0 - (1-\eta)x^*)^T (s - \eta s^0 - (1-\eta)s^*) \quad \geq \quad -\frac{\theta^2}{4}\mu. \tag{188}$$

Proof. Let us define $\tilde{w} := w - \eta w^0 - (1-\eta)w^*$. Using the definitions of $\mathcal{N}_{w^0}(\eta, \gamma, \theta)$, r, and $\mathcal{S}$, we have that

$$\begin{aligned} A\tilde{x} &= 0 \\ A^T\tilde{y} + \tilde{S} + V\tilde{z} &= 0 \\ V^T\tilde{x} + E^{-2}\tilde{z} &= E^{-1}(r_V - \eta r_V^0). \end{aligned}$$

Multiplying the second relation by $\tilde{x}^T$ on the left, and using the first and third relations along with the fact that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$, we see that

$$\begin{aligned}
\tilde{x}^T\tilde{S} &= -\tilde{x}^T V\tilde{z} \;=\; [E^{-2}\tilde{z} - E^{-1}(r_V - \eta r_V^0)]^T\tilde{z} \;=\; \|E^{-1}\tilde{z}\|^2 - (E^{-1}\tilde{z})^T(r_V - \eta r_V^0) \\
&\geq \|E^{-1}\tilde{z}\|^2 - \|E^{-1}\tilde{z}\|\|r_V - \eta r_V^0\| \;=\; \left(\|E^{-1}\tilde{z}\| - \frac{\|r_V - \eta r_V^0\|}{2}\right)^2 - \frac{\|r_V - \eta r_V^0\|^2}{4} \\
&\geq -\frac{\|r_V - \eta r_V^0\|^2}{4} \;\geq\; -\frac{\theta^2}{4}\mu.
\end{aligned}$$

■

Lemma 4.4.2 *Suppose that $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ such that $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Then, for any $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$ with $\eta \in [0, 1]$, $\gamma \in [0, 1]$ and $\theta > 0$, we have*

$$\eta(x^T s^0 + s^T x^0) \leq \left(3n + \frac{\theta^2}{4}\right)\mu. \tag{189}$$

Proof. Using the fact $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$ and (188), we obtain

$$\begin{aligned}
&x^T s - \eta(x^T s^0 + s^T x^0) + \eta^2 x^{0^T} s^0 - (1-\eta)(x^T s^* + s^T x^*) \\
&\quad +\eta(1-\eta)(x^{*^T} s^0 + s^{*^T} x^0) + (1-\eta)^2 x^{*^T} s^* \quad \geq \quad -\frac{\theta^2}{4}\mu.
\end{aligned}$$

Rearranging the terms in this equation, and using the facts that $\eta \leq x^T s / x^{0^T} s^0$, $x^{*^T} s^* = 0$, $(x, s) \geq 0$, $(x^*, s^*) \geq 0$, $(x^0, s^0) > 0$, $\eta \in [0, 1]$, $x^* \leq x^0$, and $s^* \leq s^0$, we conclude that

$$\begin{aligned}
\eta(x^T s^0 + s^T x^0) &\leq \eta^2 x^{0^T} s^0 + x^T s + \eta(1-\eta)(x^{*^T} s^0 + s^{*^T} x^0) + \frac{\theta^2}{4}\mu \\
&\leq \eta^2 x^{0^T} s^0 + x^T s + 2\eta(1-\eta) x^{0^T} s^0 + \frac{\theta^2}{4}\mu \\
&\leq 2\eta x^{0^T} s^0 + x^T s + \frac{\theta^2}{4}\mu \\
&\leq 3x^T s + \frac{\theta^2}{4}\mu \;=\; \left(3n + \frac{\theta^2}{4}\right)\mu.
\end{aligned}$$

■

Lemma 4.4.3 *Suppose $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$, $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$ for some $\eta \in [0, 1]$, $\gamma \in [0, 1]$ and $\theta > 0$, and $\bar{w}$ satisfies (184). Let W, v and u^0 be given by (175) and (185), respectively.*

Then,

$$Wu^0 - v \;=\; \tilde{T}\tilde{A}\begin{pmatrix} -x + \sigma\mu S^{-1}e + \eta(x^0 - \bar{x}) + \eta D^2(s^0 - \bar{s}) \\ E^{-1}(r_V - \eta r_V^0) \end{pmatrix}. \tag{190}$$

Proof. Using the fact that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$ along with (153), (168) and (184), we easily obtain that

$$\begin{aligned} \begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} &= \begin{pmatrix} \eta r_p^0 \\ \eta E^{-1} r_V^0 + E^{-1}(r_V - \eta r_V^0) \end{pmatrix} \\ &= \eta\tilde{A}\begin{pmatrix} x^0 - \bar{x} \\ E^{-2}(z^0 - \bar{z}) \end{pmatrix} + \tilde{A}\begin{pmatrix} 0 \\ E^{-1}(r_V - \eta r_V^0) \end{pmatrix}, \qquad (191) \\ s^0 - \bar{s} &= -A^T(y^0 - \bar{y}) - V(z^0 - \bar{z}) + r_d^0. \qquad (192) \end{aligned}$$

Using relations (152), (153), (175), (168), (185), (191) and (192), we obtain

$$\begin{aligned} Wu^0 - v &= \tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T u^0 - \tilde{T}\tilde{A}\begin{pmatrix} x - \sigma\mu S^{-1}e - D^2 r_d \\ 0 \end{pmatrix} + \tilde{T}\begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\ &= -\eta\tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\begin{pmatrix} y^0 - \bar{y} \\ z^0 - \bar{z} \end{pmatrix} - \tilde{T}\tilde{A}\begin{pmatrix} x - \sigma\mu S^{-1}e - \eta D^2 r_d^0 \\ 0 \end{pmatrix} + \tilde{T}\begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix} \\ &= -\eta\tilde{T}\tilde{A}\begin{pmatrix} D^2A^T(y^0 - \bar{y}) + D^2V(z^0 - \bar{z}) - D^2 r_d^0 \\ E^{-2}(z^0 - \bar{z}) \end{pmatrix} \\ &\quad - \tilde{T}\tilde{A}\begin{pmatrix} x - \sigma\mu S^{-1}e \\ 0 \end{pmatrix} + \tilde{T}\begin{pmatrix} r_p \\ E^{-1}r_V \end{pmatrix}, \\ &= -\eta\tilde{T}\tilde{A}\begin{pmatrix} -D^2(s^0 - \bar{s}) \\ E^{-2}(z^0 - \bar{z}) \end{pmatrix} - \tilde{T}\tilde{A}\begin{pmatrix} x - \sigma\mu S^{-1}e \\ 0 \end{pmatrix} \\ &\quad + \eta\tilde{T}\tilde{A}\begin{pmatrix} x^0 - \bar{x} \\ E^{-2}(z^0 - \bar{z}) \end{pmatrix} + \tilde{T}\tilde{A}\begin{pmatrix} 0 \\ E^{-1}(r_V - \eta r_V^0) \end{pmatrix}, \end{aligned}$$

which yields equation (190), as desired. ∎

We now turn to the proof of Lemma 4.3.4.

Proof of Lemma 4.3.4: Since $w \in \mathcal{N}_{w^0}(\gamma,\theta)$, we have that $x_i s_i \geq (1-\gamma)\mu$ for all i, which implies

$$\|(XS)^{-1/2}\| \;\leq\; \frac{1}{\sqrt{(1-\gamma)\mu}}. \tag{193}$$

Note that $\|Xs - \sigma\mu e\|$, when viewed as a function of $\sigma \in [0,1]$, is convex. Hence, it is maximized at one of its endpoints, which, together with the facts $\|Xs-\mu e\| < \|Xs\|$ and $\sigma \in [\underline{\sigma},\overline{\sigma}] \subset [0,1]$, immediately implies that

$$\|Xs - \sigma\mu e\| \;\leq\; \|Xs\| \;\leq\; \|Xs\|_1 \;=\; x^T s \;=\; n\mu. \tag{194}$$

Using the fact that $(x^0,s^0) \geq |(\bar{x},\bar{s})|$ together with Lemma 4.4.2, we obtain that

$$\begin{aligned}\eta\|S(x^0-\bar{x})+X(s^0-\bar{s})\| &\leq \eta\{\|S(x^0-\bar{x})\|+\|X(s^0-\bar{s})\|\} \;\leq\; 2\eta\{\|Sx^0\|+\|Xs^0\|\}\\ &\leq 2\eta(x^Ts^0+x^Ts^0) \;\leq\; \left(6n+\frac{\theta^2}{2}\right)\mu. \end{aligned}\tag{195}$$

Since $w \in \mathcal{N}_{w^0}(\gamma,\theta)$, there exists $\eta \in [0,1]$ such that $w \in \mathcal{N}_{w^0}(\eta,\gamma,\theta)$. It is clear that the requirements of Lemma 4.4.3 are met, so equation (190) holds. By (151), (152) and (190), we see that

$$\begin{aligned}\|v-Wu^0\| &= \left\|\tilde{T}\tilde{A}\tilde{D}\begin{pmatrix}(XS)^{-1/2}\{Xs-\sigma\mu e-\eta[S(x^0-\bar{x})+X(s^0-\bar{s})]\}\\ r_V-\eta r_V^0\end{pmatrix}\right\|\\ &\leq \|\tilde{T}\tilde{A}\tilde{D}\|\Big\{\|(XS)^{-1/2}\|\Big[\|Xs-\sigma\mu e\|+\eta\|X(s^0-\bar{s})+S(x^0-\bar{x})\|\Big]\\ &\quad +\|r_V-\eta r_V^0\|\Big\},\\ &\leq \varphi_{\tilde{A}}\left\{\frac{1}{\sqrt{(1-\gamma)\mu}}\left[n\mu+\left(6n+\frac{\theta^2}{2}\right)\mu\right]+\theta\sqrt{\mu}\right\} \;=\; \Psi\sqrt{\mu},\end{aligned}$$

where the last inequality follows from Proposition 4.3.1, relations (193), (194), (195), and the assumption that $w \in \mathcal{N}_{w^0}(\gamma,\theta)$. ∎

4.4.2 "Outer" Iteration Results – Proof of Theorem 4.2.1

In this subsection, we will present the proof of Theorem 4.2.1. Specifically, we will show that the Inexact PDIPF Algorithm obtains an ϵ-approximate solution to (137)–(140) in $\mathcal{O}(n^2\log(1/\epsilon))$ outer iterations.

Throughout this section, we use the following notation:

$$w(\alpha) := w + \alpha\Delta w, \quad \mu(\alpha) := \mu(w(\alpha)), \quad r(\alpha) := r(w(\alpha)).$$

Lemma 4.4.4 *Assume that Δw satisfies (164)-(167) for some $\sigma \in \Re$, $w \in \Re^{2n+m+l}$ and $(p, q) \in \Re^n \times \Re^l$. Then, for every $\alpha \in \Re$, we have:*

(a) $X(\alpha)s(\alpha) = (1-\alpha)Xs + \alpha\sigma\mu e - \alpha p + \alpha^2\Delta X\Delta s$;

(b) $\mu(\alpha) = [1-\alpha(1-\sigma)]\mu - \alpha p^T e/n + \alpha^2\Delta x^T\Delta s/n$;

(c) $(r_p(\alpha), r_d(\alpha)) = (1-\alpha)(r_p, r_d)$;

(d) $r_V(\alpha) = (1-\alpha)r_V + \alpha q$.

Proof. Using (166), we obtain

$$\begin{aligned} X(\alpha)s(\alpha) &= (X + \alpha\Delta X)(s + \alpha\Delta s) \\ &= Xs + \alpha(X\Delta s + S\Delta x) + \alpha^2\Delta X\Delta s \\ &= Xs + \alpha(-Xs + \sigma\mu e - p) + \alpha^2\Delta X\Delta s \\ &= (1-\alpha)Xs + \alpha\sigma\mu e - \alpha p + \alpha^2\Delta X\Delta s, \end{aligned}$$

thereby showing that (a) holds. Left multiplying the above equality by e^T and dividing the resulting expression by n, we easily conclude that (b) holds. Statement (c) can be easily verified by means of (164) and (165), while statement (d) follows from (167). ∎

Lemma 4.4.5 *Assume that Δw satisfies (164)-(167) for some $\sigma \in \Re$, $w \in \Re^{2n}_{++} \times \Re^{m+l}$ and $(p, q) \in \Re^n \times \Re^l$ such that $\|p\|_\infty \leq \gamma\sigma\mu/4$. Then, for every scalar α satisfying*

$$0 \leq \alpha \leq \min\left\{1\,,\ \frac{\sigma\mu}{4\,\|\Delta X\Delta s\|_\infty}\right\}, \tag{196}$$

we have

$$\frac{\mu(\alpha)}{\mu} \ \geq\ 1-\alpha. \tag{197}$$

Proof. Since $\|p\|_\infty \leq \gamma\sigma\mu/4$, we easily see that

$$|p^T e/n| \;\leq\; \|p\|_\infty \;\leq\; \sigma\mu/4. \tag{198}$$

Using this result and Lemma 4.4.4(b), we conclude for every α satisfying (196) that

$$\begin{aligned}
\mu(\alpha) &= [1-\alpha(1-\sigma)]\mu - \alpha p^T e/n + \alpha^2 \Delta x^T \Delta s/n \\
&\geq [1-\alpha(1-\sigma)]\mu - \frac{1}{4}\,\alpha\sigma\mu + \alpha^2 \Delta x^T \Delta s/n \\
&\geq (1-\alpha)\mu + \frac{1}{4}\,\alpha\sigma\mu - \alpha^2 \|\Delta X \Delta s\|_\infty \\
&\geq (1-\alpha)\mu.
\end{aligned}$$

■

Lemma 4.4.6 *Assume that Δw is a (τ_p, τ_q)-search direction, where τ_p and τ_q are given by (170) and (171), respectively. Assume also that $\sigma > 0$ and that $w \in \mathcal{N}_{w^0}(\gamma, \theta)$ with $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$, $\gamma \in [0,1]$ and $\theta \geq 0$. Then, $w(\alpha) \in \mathcal{N}_{w^0}(\gamma, \theta)$ for every scalar α satisfying*

$$0 \leq \alpha \leq \min\left\{1\,,\; \frac{\gamma\sigma\mu}{4\,\|\Delta X \Delta s\|_\infty}\right\}. \tag{199}$$

Proof. Since $w \in \mathcal{N}_{w^0}(\gamma, \theta)$, there exists $\eta \in [0,1]$ such that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$. We will show that $w(\alpha) \in \mathcal{N}_{w^0}((1-\alpha)\eta, \gamma, \theta) \subseteq \mathcal{N}_{w^0}(\gamma, \theta)$ for every α satisfying (199).

First, we note that $(r_p(\alpha), r_d(\alpha)) = (1-\alpha)\eta(r_p^0, r_d^0)$ by Lemma 4.4.4(c) and the definition of $\mathcal{N}_{w^0}(\eta, \gamma, \theta)$. Next, it follows from Lemma 4.4.5 that (197) holds for every α satisfying (196), and hence (199) due to $\gamma \in [0,1]$. Thus, for every α satisfying (199), we have

$$(1-\alpha)\eta \leq \;\frac{\mu(\alpha)}{\mu}\eta \;\leq\; \frac{\mu(\alpha)}{\mu}\,\frac{\mu}{\mu_0} \;=\; \frac{\mu(\alpha)}{\mu_0}. \tag{200}$$

Now, it is easy to see that for every $u \in \Re^n$ and $\tau \in [0,n]$, there holds $\|u - \tau(u^T e/n)e\|_\infty \leq (1+\tau)\|u\|_\infty$. Using this inequality twice, the fact that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$, relation (170) and statements (a) and (b) of Lemma 4.4.4, we conclude for every α satisfying (199) that

$$X(\alpha)s(\alpha) - (1-\gamma)\,\mu(\alpha)e$$

$$
\begin{aligned}
&= (1-\alpha)\left[Xs-(1-\gamma)\mu e\right]+\alpha\gamma\sigma\mu e-\alpha\left[p-(1-\gamma)\left(\frac{p^Te}{n}\right)e\right]\\
&\quad+\alpha^2\left[\Delta X\Delta s-(1-\gamma)\left(\frac{\Delta x^T\Delta s}{n}\right)e\right]\\
&\geq \alpha\left[\gamma\sigma\mu-\left\|p-(1-\gamma)\frac{p^Te}{n}\,e\right\|_\infty-\alpha\left\|\Delta X\Delta s-(1-\gamma)\frac{\Delta x^T\Delta s}{n}\,e\right\|_\infty\right]e\\
&\geq \alpha\left(\gamma\sigma\mu-2\|p\|_\infty-2\alpha\|\Delta X\Delta s\|_\infty\right)e \;\geq\; \alpha\left(\gamma\sigma\mu-\frac{1}{2}\gamma\sigma\mu-\frac{1}{2}\gamma\sigma\mu\right)e \;=\; 0.
\end{aligned}
$$

Next, by Lemma 4.4.4(d), we have that

$$
r_V(\alpha)=(1-\alpha)r_V+\alpha q=\;(1-\alpha)\eta r_V^0+\hat{a},
$$

where $\hat{a}=(1-\alpha)(r_V-\eta r_V^0)+\alpha q$. To complete the proof, it suffices to show that $\|\hat{a}\|\leq\theta\sqrt{\mu(\alpha)}$ for every α satisfying (199). By using equation (171) and Lemma 4.4.4(b) along with the facts that $\|r_V-\eta r_V^0\|\leq\theta\sqrt{\mu}$ and $\alpha\in[0,1]$, we have

$$
\begin{aligned}
\|\hat{a}\|^2-\theta^2\mu(\alpha) &= (1-\alpha)^2\|r_V-\eta r_V^0\|^2+2\alpha(1-\alpha)[r_V-\eta r_V^0]^Tq+\alpha^2\|q\|^2-\theta^2\mu(\alpha)\\
&\leq (1-\alpha)^2\theta^2\mu+2\alpha(1-\alpha)\theta\sqrt{\mu}\|q\|+\alpha^2\|q\|^2\\
&\quad-\theta^2\left\{[1-\alpha(1-\sigma)]\mu-\alpha\frac{p^Te}{n}+\alpha^2\frac{\Delta x^T\Delta s}{n}\right\}\\
&\leq \alpha^2\|q\|^2+2\alpha\theta\sqrt{\mu}\|q\|-\alpha\theta^2\sigma\mu+\alpha\theta^2\frac{p^Te}{n}-\alpha^2\theta^2\frac{\Delta x^T\Delta s}{n}\\
&\leq \alpha\left[\|q\|^2+2\theta\sqrt{\mu}\|q\|-\left(1-\frac{\gamma}{4}\right)\theta^2\sigma\mu+\theta^2\alpha\|\Delta X\Delta s\|_\infty\right]\\
&\leq \alpha\left[\|q\|^2+2\theta\sqrt{\mu}\|q\|-\left(1-\frac{\gamma}{2}\right)\theta^2\sigma\mu\right]\;\leq\;0,
\end{aligned}
$$

where the last inequality follows from the quadratic formula and the fact that $\|q\|\leq\tau_q$, where τ_q is given by (171). ∎

Next, we derive a lower bound on the stepsize of the Inexact PDIPF Algorithm.

Lemma 4.4.7 *In every iteration of the Inexact PDIPF Algorithm, the step length $\bar{\alpha}$ satisfies*

$$
\bar{\alpha}\;\geq\;\min\left\{1,\frac{\min\{\gamma\sigma,1-\frac{5}{4}\sigma\}\mu}{4\,\|\Delta X\Delta s\|_\infty}\right\} \tag{201}
$$

and

$$
\mu(\bar{\alpha})\;\leq\;\left[1-\left(1-\frac{5}{4}\,\sigma\right)\frac{\bar{\alpha}}{2}\right]\mu. \tag{202}
$$

Proof. We know that Δw is a (τ_p, τ_q)-search direction in every iteration of the Inexact PDIPF Algorithm, where τ_p and τ_q are given by (170) and (171). Hence, by Lemma 4.4.6, the quantity $\tilde{\alpha}$ computed in step (g) of the Inexact PDIPF Algorithm satisfies

$$\tilde{\alpha} \;\geq\; \min\left\{1, \frac{\gamma\sigma\mu}{4\,\|\Delta X \Delta s\|_\infty}\right\}. \tag{203}$$

Moreover, by (198), it follows that the coefficient of α in the expression for $\mu(\alpha)$ in Lemma 4.4.4(b) satisfies

$$-(1-\sigma)\mu - \frac{p^T e}{n} \;\leq\; -(1-\sigma)\mu + \|p\|_\infty \;\leq\; -(1-\sigma)\mu + \frac{1}{4}\gamma\sigma\mu \;=\; -\left(1 - \frac{5}{4}\sigma\right)\mu \;<\; 0, \tag{204}$$

since $\sigma \in (0, 4/5)$. Hence, if $\Delta x^T \Delta s \leq 0$, it is easy to see that $\bar{\alpha} = \tilde{\alpha}$, and hence that (201) holds in view of (203). Moreover, by Lemma 4.4.4(b) and (204), we have

$$\mu(\bar{\alpha}) \;\leq\; [1 - \bar{\alpha}(1-\sigma)]\mu - \bar{\alpha}\frac{p^T e}{n} \;\leq\; \left[1 - \left(1 - \frac{5}{4}\sigma\right)\bar{\alpha}\right]\mu \;\leq\; \left[1 - \left(1 - \frac{5}{4}\sigma\right)\frac{\bar{\alpha}}{2}\right]\mu,$$

showing that (202) also holds. We now consider the case where $\Delta x^T \Delta s > 0$. In this case, we have $\bar{\alpha} = \min\{\alpha_{\min}\,,\, \tilde{\alpha}\}$, where $\alpha_{\min}$ is the unconstrained minimum of $\mu(\alpha)$. It is easy to see that

$$\alpha_{\min} \;=\; \frac{n\mu(1-\sigma) + p^T e}{2\Delta x^T \Delta s} \;\geq\; \frac{n[\mu(1-\sigma) - \frac{1}{4}\sigma\mu]}{2\Delta x^T \Delta s} \;\geq\; \frac{\mu(1 - \frac{5}{4}\sigma)}{2\,\|\Delta X \Delta s\|_\infty}.$$

The last two observations together with (203) imply that (201) holds in this case too. Moreover, since the function $\mu(\alpha)$ is convex, it must lie below the function $\phi(\alpha)$ over the interval $[0, \alpha_{\min}]$, where $\phi(\alpha)$ is the affine function interpolating $\mu(\alpha)$ at $\alpha = 0$ and $\alpha = \alpha_{\min}$. Hence,

$$\mu(\bar{\alpha}) \;\leq\; \phi(\bar{\alpha}) \;=\; [1 - (1-\sigma)\frac{\bar{\alpha}}{2}]\mu - \bar{\alpha}\frac{p^T e}{2n} \;\leq\; \left[1 - \left(1 - \frac{5}{4}\sigma\right)\frac{\bar{\alpha}}{2}\right]\mu, \tag{205}$$

where the second inequality follows from (204). We have thus shown that $\bar{\alpha}$ satisfies (202).

■

Our next task will be to show that the stepsize $\bar{\alpha}$ remains bounded away from zero. In view of (201), it suffices to show that the quantity $\|\Delta X \Delta s\|_\infty$ can be suitably bounded. The next lemma addresses this issue.

Lemma 4.4.8 *Let $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ be such that $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$, and let $w \in \mathcal{N}_{w^0}(\gamma, \theta)$ for some $\gamma \geq 0$ and $\theta \geq 0$. Then, the inexact search direction Δw used in the Inexact PDIPF Algorithm satisfies*

$$\begin{aligned} \max(\|D^{-1}\Delta x\|, \|D\Delta s\|) &\leq \left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)^{\frac{1}{2}} \sqrt{n\mu} \\ &\quad + \frac{1}{\sqrt{1-\gamma}} \left(\frac{\gamma\sigma}{4}\sqrt{n} + 6n + \frac{\theta^2}{2}\right) \sqrt{\mu} + \theta\sqrt{\mu}. \end{aligned} \tag{206}$$

Proof. Since $w \in \mathcal{N}_{w^0}(\gamma, \theta)$, there exists $\eta \in [0, 1]$ such that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$. Let $\widetilde{\Delta w} := \Delta w + \eta(w^0 - w^*)$. Using relations (164), (165), (167), and the fact that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$, we easily see that

$$A\widetilde{\Delta x} = 0 \tag{207}$$

$$A^T\widetilde{\Delta y} + \widetilde{\Delta s} + V\widetilde{\Delta z} = 0, \tag{208}$$

$$V^T\widetilde{\Delta x} + E^{-2}\widetilde{\Delta z} = E^{-1}(q - r_V + \eta r_V^0). \tag{209}$$

Pre-multiplying (208) by $\widetilde{\Delta x}^T$ and using (207) and (209), we obtain

$$\begin{aligned} \widetilde{\Delta x}^T\widetilde{\Delta s} &= -\widetilde{\Delta x}^T V\widetilde{\Delta z} = [E^{-2}\widetilde{\Delta z} - E^{-1}(q - r_V + \eta r_V^0)]^T\widetilde{\Delta z} \\ &= \|E^{-1}\widetilde{\Delta z}\|^2 - (q - r_V + \eta r_V^0)^T(E^{-1}\widetilde{\Delta z}) \\ &\geq \|E^{-1}\widetilde{\Delta z}\|^2 - \|q - r_V + \eta r_V^0\| \, \|E^{-1}\widetilde{\Delta z}\| \geq -\frac{\|q - r_V + \eta r_V^0\|^2}{4}. \end{aligned} \tag{210}$$

Next, we multiply equation (166) by $(XS)^{-1/2}$ to obtain $D^{-1}\Delta x + D\Delta s = H(\sigma) - (XS)^{-1/2}p$, where $H(\sigma) := -(XS)^{1/2}e + \sigma\mu(XS)^{-1/2}e$. Equivalently, we have that

$$D^{-1}\widetilde{\Delta x} + D\widetilde{\Delta s} = H(\sigma) - (XS)^{-1/2}p + \eta\left[D(s^0 - s^*) + D^{-1}(x^0 - x^*)\right] =: g.$$

Taking the squared norm of both sides of the above equation and using (210), we obtain

$$\begin{aligned} \|D^{-1}\widetilde{\Delta x}\|^2 + \|D\widetilde{\Delta s}\|^2 &= \|g\|^2 - 2\widetilde{\Delta x}^T\widetilde{\Delta s} \leq \|g\|^2 + \frac{\|q - r_V + \eta r_V^0\|^2}{2} \\ &\leq \left(\|g\| + \frac{\|q\| + \|r_V - \eta r_V^0\|}{\sqrt{2}}\right)^2 \leq \left(\|g\| + \theta\sqrt{\mu}\right)^2, \end{aligned}$$

since $\|q\| + \|r_V - \eta r_V^0\| \leq \left[\sqrt{2} - 1\right] \theta\sqrt{\mu} + \theta\sqrt{\mu} = \sqrt{2}\theta\sqrt{\mu}$ by (168), (171), and the fact that $1 + (1 - \gamma/2)\sigma \leq 2$. Thus, we have

$$\begin{aligned}\max(\|D^{-1}\widetilde{\Delta x}\|\,,\, \|D\widetilde{\Delta s}\|) \;&\leq\; \|g\| + \theta\sqrt{\mu} \\ &\leq\; \|H(\sigma)\| + \|(XS)^{-1/2}\|\,\|p\| + \eta\left[\|D(s^0 - s^*)\| + \|D^{-1}(x^0 - x^*)\|\right] + \theta\sqrt{\mu}.\end{aligned}$$

This, together with the triangle inequality, the definitions of D and $\widetilde{\Delta w}$, and the fact that $w \in \mathcal{N}_{w^0}(\eta, \gamma, \theta)$, imply that

$$\begin{aligned}&\max(\|D^{-1}\Delta x\|, \|D\Delta s\|) \\ &\leq \;\; \|H(\sigma)\| + \|(XS)^{-1/2}\|\,\|p\| + 2\eta\left[\|D(s^0 - s^*)\| + \|D^{-1}(x^0 - x^*)\|\right] + \theta\sqrt{\mu} \\ &\leq \;\; \|H(\sigma)\| + \|(XS)^{-1/2}\|\,\|p\| + 2\eta\|(XS)^{-1/2}\|\left[\|X(s^0 - s^*)\| + \|S(x^0 - x^*)\|\right] + \theta\sqrt{\mu} \\ &\leq \;\; \|H(\sigma)\| + \frac{1}{\sqrt{(1-\gamma)\mu}}\left[\|p\| + 2\eta\left(\|X(s^0 - s^*)\| + \|S(x^0 - x^*)\|\right)\right] + \theta\sqrt{\mu}. \qquad (211)\end{aligned}$$

It is well-known (see e.g., [61]) that

$$\|H(\sigma)\| \leq \;\; \left(1 - 2\sigma + \tfrac{\sigma^2}{1-\gamma}\right)^{\frac{1}{2}}\sqrt{n\mu}. \qquad (212)$$

Moreover, using the fact that $s^* \leq s^0$ and $x^* \leq x^0$ along with Lemma 4.4.2, we obtain

$$\eta\left(\|X(s^0 - s^*)\| + \|S(x^0 - x^*)\|\right) \;\leq\; \eta(s^T x^0 + x^T s^0) \;\leq\; \left(3n + \frac{\theta^2}{4}\right)\mu. \qquad (213)$$

The result now follows by noting that $\|p\| \leq \sqrt{n}\|p\|_\infty$, and by incorporating inequalities (212), (213) and (170) into (211). ∎

We are now ready to prove Theorem 4.2.1.

Proof of Theorem 4.2.1: Let Δw^k denote the search direction, and let $r^k = r(w^k)$ and $\mu_k = \mu(w^k)$, at the k-th iteration of the Inexact PDIPF Algorithm. Clearly, $w^k \in \mathcal{N}_{w^0}(\gamma, \theta)$. Hence, using Lemma 4.4.8, assumption (172) and the inequality

$$\|\Delta X^k \Delta s^k\|_\infty \;\leq\; \|\Delta X^k \Delta s^k\| \;\leq\; \|(D^k)^{-1}\Delta x^k\|\,\|D^k \Delta s^k\|, \qquad (214)$$

we easily see that $\|\Delta X^k \Delta s^k\|_\infty = \mathcal{O}(n^2)\mu_k$. Using this conclusion together with assumption (172) and Lemma 4.4.7, we see that, for some universal constant $\beta > 0$, we have

$$\mu_{k+1} \;\leq\; \left(1 - \frac{\beta}{n^2}\right)\mu_k, \quad \forall k \geq 0.$$

Using this observation and some standard arguments (see, for example, Theorem 3.2 of [130]), we easily see that the Inexact PDIPF Algorithm generates an iterate $w^k \in \mathcal{N}_{w^0}(\gamma, \theta)$ satisfying $\mu_k/\mu_0 \leq \epsilon$ within $\mathcal{O}\left(n^2 \log(1/\epsilon)\right)$ iterations. The theorem now follows from this conclusion and the definition of $\mathcal{N}_{w^0}(\gamma, \theta)$. ∎

4.5 Concluding Remarks

We have shown that the long-step PDIPF algorithm for LP based on an iterative linear solver presented in [80] can be extended to the context of CQP. This was not immediately obvious at first since the standard normal equation for CQP does not fit into the mold required for the results of [81] to hold. By considering the ANE, we were able to use the results about the MWB preconditioner developed in [81] in the context of CQP. Another difficulty we encountered was the proper choice of the starting iterate u^0 for the iterative linear solver. By choosing $u^0 = 0$ as in the LP case, we obtain $\|v - Wu^0\| = \|v\|$, which can only be shown to be $\mathcal{O}(\max\{\mu, \sqrt{\mu}\})$. In this case, for every $\mu > 1$, Proposition 4.3.2 would guarantee that the number of inner iterations of the iterative linear solver is

$$\mathcal{O}\left(\psi(\varphi_{\tilde{A}}^2) \max\left\{ \log\left(c(\varphi_{\tilde{A}}^2) n \varphi_{\tilde{A}}\right) , \log\mu \right\}\right),$$

a bound which depends on the logarithm of the current duality gap. On the other hand, Theorem 4.3.5 shows that choosing u^0 as in (185) results in a bound that does not depend on the current duality gap.

We observe that under exact arithmetic, the CG algorithm applied to $Wu = v$ generates an exact solution in at most $m + l$ iterations (since $W \in \Re^{(m+l)\times(m+l)}$). It is clear, then, that the bound (187) is generally worse than the well-known finite termination bound for CG. However, our results in Section 4.3 were given for a family of iterative linear solvers, only one member of which is CG. Also, under finite precision arithmetic, the CG algorithm loses its finite termination property, and its convergence rate behavior in this case is still an active topic of research (see e.g., [45]). Certainly, the impact of finite precision arithmetic on our results is an interesting open issue.

Clearly, the MWB preconditioner is not suitable for dense CQP problems since, in this

case, the cost to construct the MWB is comparable to the cost to form and factorize $\tilde{A}\tilde{D}^2\tilde{A}^T$, and each inner iteration would require $\Theta((m+l)^2)$ arithmetic operations, the same cost as a forward and back substitution. There are, however, some classes of CQP problems for which the method proposed in this chapter might be useful. One class of problems for which PDIPF methods based on MWB preconditioners might be useful are those for which bases of $\tilde{A}$ are sparse but the ANE coefficient matrices $\tilde{A}\tilde{D}^2\tilde{A}^T$ are dense; this situation generally occurs in sparse CQP problems for which n is much larger than $m+l$. Other classes of problems for which our method might be useful are network flow problems. The paper [111] developed interior-point methods for solving the minimum cost network flow problem based on iterative linear solvers with maximum spanning tree preconditioners. Related to this work, we believe that the following two issues could be investigated: (i) will the incorporation of the correction term p defined in (161) in the algorithm implemented in [111] improve the convergence of the method? (ii) whether our algorithm might be efficient for network flow problems where the costs associated with the arcs are quadratic functions of the arc flows? Identification of other classes of CQP problems which could be efficiently solved by the method proposed in this chapter is another topic for future research.

Regarding the second question above, it is easy to see (after a suitable permutation of the variables) that $V^T = \begin{pmatrix} I & 0 \end{pmatrix}$ and E^2 is a positive diagonal matrix whose diagonal elements are the positive quadratic coefficients. In this case, it can be shown that $\tilde{A}$ is totally unimodular, hence $\varphi^2_{\tilde{A}} \leq (m+l)(n-m+1)$ by Cramer's Rule (see [81]).

The usual way of defining the dual residual is as the quantity

$$R_d := A^T y + s - VE^2V^T x - c,$$

which, in view of (143) and (144), can be written in terms of the residuals r_d and r_V as

$$R_d = r_d - VEr_V. \tag{215}$$

Note that, along the iterates generated by the Inexact PDIPF Algorithm, we have $r_d = \mathcal{O}(\mu)$ and $r_V = \mathcal{O}(\sqrt{\mu})$, implying that $R_d = \mathcal{O}(\sqrt{\mu})$. Hence, while the usual primal residual converges to 0 according to $\mathcal{O}(\mu)$, the usual dual residual does so according to $\mathcal{O}(\sqrt{\mu})$. This is a unique feature of the convergence analysis of our algorithm in that it contrasts with the

analysis of other interior-point PDIPF algorithms, where the primal and dual residuals are required to go to zero at the same rate. The convergence analysis under these circumstances is possible due to the specific form of the $\mathcal{O}(\sqrt{\mu})$-term present in (215), i.e., one that lies in the range space of VE.

CQP problems where V is explicitly available arise frequently in the literature. One important example arises in portfolio optimization (see [21]), where the rank of V is often small. In such problems, l represents the number of observation periods used to estimate the data for the problem. We believe that the Inexact PDIPF Algorithm could be of particular use for this type of application.

CHAPTER V

A MODIFIED NEARLY EXACT METHOD FOR SOLVING A LOW-RANK TRUST REGION SUBPROBLEM

5.1 Preliminary Remarks

Trust region algorithms are classical methods for solving both convex and nonconvex nonlinear optimization problems. They are known to possess strong convergence properties (see Fletcher [32]). At each iteration of a trust region method, the following is specified: i) a "simple" approximation $\phi(\tilde{x})$ to the objective function, called the model, and; ii) a region T around the current iterate x, where $\phi(\tilde{x})$ is believed to provide a good approximation to the objective function. An approximate solution p of the subproblem $\min_p\{\phi(x+p) : x+p \in T\}$ is then computed, and the next iterate $\hat{x}$ is set to be $\hat{x} := x + p$ provided there is a "significant" objective function progress; otherwise, we define the next iterate $\hat{x}$ as $\hat{x} := x$. In both cases, the region T might be updated and the process is then repeated until a desirable iterate is obtained.

In most trust region methods, the above subproblem is either of or reduces to the following form:

$$\text{minimize}\,\{\, q(p) : \|p\|_M \leq \Delta \,\} \tag{216}$$

where Δ is a positive parameter, M is a symmetric positive-definite matrix referred to as the scaling matrix, $\|\cdot\|_M$ is the *M-norm* defined as

$$\|x\|_M = \sqrt{x^T M x}, \quad \forall x \in \Re^n,$$

and $q : \Re^n \to \Re$ is the quadratic function defined as

$$q(p) = g^T p + \frac{1}{2} p^T H p, \quad \forall p \in \Re^n, \tag{217}$$

for some $g \in \Re^n$ and symmetric matrix $H \in \Re^{n\times n}$. The matrix H can be either the Hessian of the objective function or some approximation of it.

There are at least three well-known methods available in the literature for finding an "approximate" solution of TR subproblem (216), which achieves at least as much reduction in the model q as the reduction achieved by the so called "Cauchy point" (see for example Moré [88] and Chapter 4 of Nocedal and Wright [98]). The first method is the *dogleg* method proposed by Powell [105], and later modified by Dennis and Mei [29], which is appropriate when the model Hessian H is positive definite. Recently, Zhang and Xu [131] proposed a *dogleg* method for the case when H is indefinite, which is based on the estimation of the most negative eigenvalue of H and computation of the stable Bunch-Parlett factorization of H (see [18]). The second method is the *two-dimensional subspace minimization* method proposed by Shultz et al. [39], which can be applied when H is indefinite, though it also requires an estimate of the most negative eigenvalue of H. The third method, due to Steihaug [118], is most appropriate when H is the Hessian of the objective function and when this matrix is large and sparse.

Besides the three "approximate" methods mentioned above, there is a method due to Moré and Sorensen [89], which finds an approximate solution of the TR subproblem (216) in a stronger sense (see (224)). Following standard convention, we will refer to such solutions as "nearly exact" (NE) solutions and to the methods for computing them as NE methods. Since the NE method of [89] requires repeated computations of Cholesky factorizations of diagonal displacements of H, it is suitable only for small- to medium-sized problems.

The main goal of this chapter is to develop a method for computing NE solutions of the TR subproblem (216) when H and M are large-scale matrices having the following special structures:

$$H = D + VEV^T, \tag{218}$$

$$M = \tilde{D} + \tilde{V}\tilde{E}\tilde{V}^T \succ 0, \tag{219}$$

where D, $\tilde{D}$ and $\tilde{E}$ are positive diagonal matrices, V and $\tilde{V}$ have a few columns (say less than 10), and E is a diagonal matrix. We will refer to the resulting subproblem as the "low-rank trust region" (LRTR) subproblem. We will show that every step of the NE method of [89] can be properly modified to handle the LRTR subproblem and also that the resulting

modified NE (MNE) method is quite efficient and robust for computing NE solutions of large-scale LRTR subproblems.

LRTR subproblems arise in several contexts. For example, when using trust region methods to solve unconstrained or linear-equality constrained minimization problems, the matrix H is usually obtained by using a low-rank update (memoryless) formula and the resulting H has the structure specified in (218). In such a case, the scaling matrix M is chosen as either the identity matrix or some other positive definite matrix whose structure is as specified in (219) and depends on the specific problem at hand. It is well known that many constrained minimization problems can be solved by minimizing a sequence of unconstrained ones, obtained by using either the penalty, log-barrier, augmented Lagrangian multiplier (see for example [98]), or modified log-barrier methods (see Polyak [101]). Thus, the NE method developed in this chapter for solving the LRTR subproblem can potentially be used in solving many optimization problems.

The following notations are used throughout this chapter. We denote the k-th coordinate vector by e_k and the identity matrix by I. Their dimensions are always clear from the context. The symbol $\Re^n$ denote the n-dimensional Euclidean space. The set of all $m \times n$ matrices with real entries is denoted by $\Re^{m\times n}$. For $J \subseteq \{1,\ldots,n\}$ and $w \in \Re^n$, we let w_J denotes the subvector $[w_i]_{i\in J}$; moreover, if E is an $m \times n$ matrix then E_J denotes the $m \times |J|$ submatrix of E corresponding to J. For a vector $w \in \Re^n$, Diag(w) denote the diagonal matrix whose i-th diagonal element is w_i for $i = 1,\ldots,n$, and for any real number α, w^α denote the vector whose i-th component is w_i^α whenever it is well-defined for $i = 1,\ldots,n$. The Euclidean norm, the 1-norm and the ∞-norm are denoted by $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$, respectively. For a matrix E, Im(E) denotes the subspace generated by the columns of E and Ker(E) denotes the subspace orthogonal to the rows of E. The superscript ^{T} denotes transpose. For any real symmetric matrix E, $\lambda_{\min}(E)$ (resp., $\lambda_{\max}(E)$) denotes the minimal (resp., maximal) eigenvalue of the matrix E; $E \succeq 0$ (resp., $E \succ 0$) denotes that E is positive semi-definite (resp., positive definite).

Before ending this section, we provide one example to show how the LRTR subproblem naturally arises in the context of solving linearly constrained minimization problems using

a log-barrier approach. Indeed, consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b, \\ & l \leq x \leq u, \end{aligned} \tag{220}$$

where $f(x)$ is twice continuously differentiable in $\Re^n$, $A \in \Re^{m \times n}$ has full row rank, l, $u \in \Re^n$ may have some components equal to $-\infty$ or $+\infty$, and m is small. The log-barrier approach applied to (220) consists of solving the following sequence of log-barrier subproblems parametrized by $\mu > 0$:

$$\begin{aligned} \text{minimize} \quad & \phi_\mu(x) \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{221}$$

where

$$\phi_\mu(x) = f(x) - \mu \sum_{i=1}^{n} \log(x_i - l_i) - \mu \sum_{i=1}^{n} \log(u_i - x_i).$$

Assume that x denotes the current iterate towards a (local) solution of (221). To find the next iterate, a typical TR method in this context computes the (potential) displacement p by solving the TR subproblem

$$\begin{aligned} \text{minimize} \quad & g^T p + \tfrac{1}{2} p^T H p \\ \text{subject to} \quad & Ap = 0, \\ & \|W^{-1} p\| \leq \Delta, \end{aligned} \tag{222}$$

where $W = \text{Diag}([(x-l)^{-2} + (u-x)^{-2}]^{-1/2}) \succ 0$, $g = \nabla \phi_\mu(x)$, $\Delta > 0$, and $H = D + VEV^T$ is an approximation to $\nabla^2 \phi_\mu(x)$ obtained by using a low-rank (memoryless) update formula. Thus, H has the structure as in (218), Now, let $B \subset \{1, \ldots, n\}$ be a basic index set and let N denote its complement. By permuting the columns of A, W and D, we may assume that

$$A = [A_B,\, A_N], \quad W = \text{Diag}(W_B,\, W_N), \quad D = \text{Diag}(D_B, D_N).$$

Since $Ap = 0$, we have $p_B = -A_B^{-1} A_N p_N$. Eliminating p_B from (222), we obtain the following equivalent LRTR subproblem

$$\begin{aligned} \text{minimize} \quad & \bar{g}^T p_N + \tfrac{1}{2} p_N^T \bar{H} p_N \\ \text{subject to} \quad & \|p_N\|_M \leq \Delta, \end{aligned} \tag{223}$$

where

$$\bar{g} = S^T g, \quad \bar{H} = S^T H S, \quad M = S^T W^{-2} S, \quad S = \begin{bmatrix} -A_B^{-1} A_N \\ I \end{bmatrix}.$$

Thus, we easily see that

$$\begin{aligned} M &= W_N^{-2} + (A_B^{-1} A_N)^T W_B^{-2} (A_B^{-1} A_N), \\ \bar{H} &= D_N + (A_B^{-1} A_N)^T D_B (A_B^{-1} A_N) + (S^T V) E (V^T S). \end{aligned}$$

Noting that A has low full rank and H has the structure specified in (218), we immediately see that the matrices M and $\bar{H}$ themselves have the structure as in (219) and (218), respectively. Thus, the subproblem (223) is indeed an LRTR subproblem.

The outline of this chapter is as follows. In Section 5.2, we review the NE method proposed by Moré and Sorensen [89]. In Section 5.3, we discuss how this method can be modified in order to solve large-scale LRTR subproblems efficiently. In Section 5.4, we first review the modified log-barrier (MLB) algorithm proposed by Polyak [101] and implement a specific version of this algorithm where the generated log-barrier subproblems are solved by a trust region method whose direction finding subproblems are of the LRTR type. The LRTR subproblems are then solved by our modified NE method. Section 5.4 also gives computational results of our implementation of the MLB method and its comparison with a version of LANCELOT [23] based on a collection extracted from CUTEr [43] of nonlinear programming problems with simple bound constraints.

5.2 Review the NE method for solving TR subproblem

It is well-known that the TR subproblems which arise in a TR method does not need to be solved exactly to guarantee the global convergence of the algorithm. For example, it has been shown by Moré and Sorensen [89] (see also [88]) that, under some mild conditions, good theoretical and numerical convergence results for a standard TR method can be obtained if p is chosen so that

$$q(p) \leq \tau_1 q^* \quad \text{and} \quad \|p\|_M \leq \tau_2 \Delta \tag{224}$$

for some positive constants τ_1 and τ_2, where q^* is the optimal value of TR subproblem (216). (Note that $q^* \leq 0$ and that $q^* = 0$ if and only if $g = 0$ and $H \succeq 0$.) We will refer to such vectors p as NE solutions of (216).

The NE method proposed by Moré and Sorensen [89] is a method for computing a NE solution p of (216). In this section, we review the technical results of the NE method of [89] for solving TR subproblem (216) (see [24, 35, 89, 88, 98] for more details). The computational difficulties of using the NE method for TR subproblems corresponding to large-scale optimization problems are also presented. But, in Section 5.3, we will show that the NE method can be suitably modified to overcome these difficulties if all the TR subproblems are constructed as LRTR ones.

This section is divided into five subsections. In Subsection 5.2.1, we discuss the necessary and sufficient optimality conditions for a global solution of the TR subproblem (216). In Subsection 5.2.2, we discuss some classical and easily verifiable sufficient conditions for p to be a NE solution of (216) (see for example [24] and [89]). In Subsection 5.2.3, we discuss how Newton method applied to a classical one dimensional nonlinear equation provides an estimate of the optimal Lagrange multiplier associated with the constraint of (216). Since the search for the optimal Lagrange multiplier requires the estimation of ever-improving lower bounds for it, we discuss in Subsection 5.2.4 how these bounds are normally generated. The complete NE method of [89] and its computational difficulties in the context of large-scale problems are also discussed in this subsection.

5.2.1 Characterization of the solution of TR subproblem

In this subsection, we provide optimality conditions which characterize the global solutions of subproblem (216).

The proof of the next lemma, which provides the above mentioned optimality conditions, is given in Theorem 7.4.1 on pp. 201 of Conn et al. [24]. (This result was obtained independently by Gay [40] and Sorensen [116].)

Lemma 5.2.1 *p is a global solution of TR subproblem (216) if only if $\|p\|_M \leq \Delta$ and there*

exists $\lambda \geq 0$ *such that*

$$H(\lambda)p = -g, \tag{225}$$

$$\lambda(\Delta - \|p\|_M) = 0, \tag{226}$$

$$H(\lambda) \succeq 0, \tag{227}$$

where $H(\lambda) \equiv H + \lambda M$ *for any* $\lambda \in \Re$. *Moreover, there exists a unique* $\lambda^* \geq 0$ *such that:*

i) $\lambda = \lambda^*$ *for every pair* (p, λ) *as above;*

ii) if $H(\lambda^*) \succ 0$ *then (216) has a unique global optimal solution.*

We now introduce some notation. Define

$$\lambda_1 \equiv \lambda_{\min}(M^{-1/2} H M^{-1/2}), \quad \hat{\lambda} \equiv \max(-\lambda_1, 0). \tag{228}$$

Moreover, we define

$$p(\lambda) \equiv -H(\lambda)^{-1} g, \tag{229}$$

for every $\lambda \in \Re$ for which the above inverse exists. It is well-known that, when $g \neq 0$, the function $\|p(\lambda)\|_M$ is strictly decreasing and convex on $(\hat{\lambda}, \infty)$.

We now describe how an exact solution for (216) can be computed, depending on which of the following three cases occur:

1) If there exists $\tilde{\lambda} \in (\hat{\lambda}, +\infty)$ such that $\tilde{\lambda}$ solves the equation

$$\|p(\lambda)\|_M = \Delta, \tag{230}$$

then $\lambda^* = \tilde{\lambda} > 0$ and $p(\tilde{\lambda})$ is the unique solution of (216). This case is usually referred to as the "easy" one. (Note that this case occurs if and only if $\lim_{\lambda \downarrow \hat{\lambda}} \|p(\lambda)\|_M > \Delta$.)

2) If $\lambda_1 > 0$ and (230) has no solution in $(\hat{\lambda}, +\infty)$, then $\lambda^* = \hat{\lambda} = 0$ and $p(0)$ is a solution of (216). (This case can easily be detected and usually referred to as the "interior convergence" one.)

3) If $\lambda_1 \leq 0$ and (230) has no solution in $(\hat{\lambda}, +\infty)$, then $\lambda^* = \hat{\lambda} = -\lambda_1$. Hence, there exist $0 \neq u \in \mathrm{Ker}(H + \hat{\lambda}M)$ and $\alpha^M \in \Re$ such that

$$\|p_{\mathrm{crt}} + \alpha^M u\|_M = \Delta,$$

where $p_{\mathrm{crt}} \equiv -H(\hat{\lambda})^{\dagger} g$, and the superscript † denotes the Moore-Penrose generalized inverse. Using Lemma 5.2.1, one easily sees that $p_{\mathrm{crt}} + \alpha^M u$ is a solution of (216). This case is usually referred to as the "hard" one.

Some steps in the NE algorithm for solving (216) require to test whether $\lambda > \hat{\lambda}$ or $\lambda > \lambda^*$. These two inequalities can be checked by using the following easily verifiable characterizations: i) $\lambda > \hat{\lambda}$ if and only if $\lambda > 0$ and $H(\lambda) \succ 0$; ii) $\lambda > \lambda^*$ if and only if $\lambda > \hat{\lambda}$ and $\|p(\lambda)\|_M < \Delta$.

5.2.2 Termination conditions

Among the three cases mentioned in Subsection 5.2.1, only the interior convergence case can be implemented exactly. When the other two cases occur, we can only expect to obtain an approximate solution of (216). In this subsection, we review some sufficient conditions for a vector $p \in \Re^n$ to be a NE solution of (216) when either the easy or hard case occurs.

While looking for a scalar $\lambda > \hat{\lambda}$ satisfying (230), we might simply stop when

$$| \, \|p(\lambda)\|_M - \Delta \, | \leq k_e \Delta,$$

where $k_e \in (0, 1)$ is a fixed tolerance. In this case, the following result establishes that $p(\lambda)$ is a NE solution of (216). Its proof is similar to the one given in Lemma 7.3.5 on pp. 195 of [24] (see also [89]).

Lemma 5.2.2 *If $\lambda > \hat{\lambda}$ satisfies $| \, \|p(\lambda)\|_M - \Delta \, | \leq k_e \Delta$ for some $k_e \in (0, 1)$, then, there holds*

$$q(p(\lambda)) \leq (1 - k_e)^2 q^*$$

.

The following result describes how an approximate version of the hard case yields NE solutions for (216). Its proof is similar to the one given in Lemma 7.3.6 on pp. 196 of [24] (see also [89]).

Lemma 5.2.3 *Suppose that* $\lambda > \lambda^*$, $\alpha \in \Re$ *and* $u \in \Re^n$ *such that* $\|u\|_M = 1$ *satisfy*

$$\alpha^2 u^T H(\lambda) u \leq k_h \left(p(\lambda)^T H(\lambda) p(\lambda) + \lambda \Delta^2 \right) \text{ and } \|p(\lambda) + \alpha u\|_M = \Delta, \tag{231}$$

for some $k_h \in (0,\, 1)$. *Then,* $q(p(\lambda) + \alpha u) \leq (1 - k_h) q^*$.

The main use of Lemma 5.2.3 is related with the hard case, but we emphasize that the result can also be applied in other cases. We now explain how to satisfy the conditions of Lemma 5.2.3 in the hard case. From the discussion in Subsection 5.2.1, we know that $\lambda^* = -\lambda_1$. Hence, if $\lambda - \lambda^*$ is sufficiently small, then $H(\lambda)$ is nearly singular and thus there exists a vector u such that $\|u\|_M = 1$ and $u^T H(\lambda) u$ is nearly zero. Once the pair (λ, u) is determined, a scalar α satisfying the equality in (231) can be easily obtained by solving the following problem

$$\begin{array}{ll} \text{minimize}_{\,\alpha} & \frac{1}{2}(p + \alpha u)^T H (p + \alpha u) + g^T (p + \alpha u) \\ \text{subject to} & \|p + \alpha u\|_M = \Delta. \end{array} \tag{232}$$

Using the well-known formula of α for the case when $M = I$ (see page 558 of [89]), it is easy to see that the optimal solution of (232) is given by

$$\alpha = \frac{\Delta^2 - \|p\|_M^2}{p^T M u + \text{sgn}(p^T M u)[(p^T M u)^2 + \Delta^2 - \|p\|_M^2]^{1/2}}, \tag{233}$$

where the function sgn : $\Re \to \Re$ is defined as $\text{sgn}(t) = 1$ if $t \geq 0$, and $\text{sgn}(t) = -1$ if $t < 0$. It can be easily verified that the right hand side of the inequality in (231) evaluated at a triple (λ, u, α) obtained as above stays bounded away from zero as $\lambda - \lambda^*$ approaches zero. Thus, as $\lambda - \lambda^*$ approaches zero, a triple (λ, u, α) obtained as above will eventually satisfy (231) when the hard case occurs.

The key part to find a triple (λ, u, α) satisfying (231) in the hard case is the computation of a vector $u = u_\lambda$ such that $u^T H(\lambda) u$ approaches zero as $\lambda \downarrow \lambda^*$. The NE method of [89] computes such a vector u by first computing the Cholesky factor of $H(\lambda)$, and then using

the LINPACK technique [22] (see also Appendix of [89]) for estimating its smallest singular value. On the other hand, since our approach for solving the low rank version of (216) does not rely on computation of Cholesky factorizations, it uses an entirely different approach to compute a vector u as above (see Subsection 5.3.3).

5.2.3 Newton update for λ

We have seen in the Subsection 5.2.1 that λ^* is a root of (230) in the easy case. Hence, given an approximation $\lambda > -\lambda_1$ of λ^*, it is natural to try to perform a Newton iteration at λ with respect to (230) to obtain a new approximation of λ^*. In this subsection, we describe the details of a Newton iteration applied to a reformulation of the nonlinear equation (230) and discuss its main properties.

Since the function $\|p(\lambda)\|_M$ goes to infinity as λ tends to $-\lambda_1$, it is highly nonlinear near $-\lambda_1$ (see [88] and [98]). As a result, Newton method applied directly to (230) might not work well when λ is near $-\lambda_1$. Reinsch [109] and Hebden [53] independently observed that Newton method applied to the following alternative reformulation of (230) works better in practice:

$$\phi(\lambda) \equiv \frac{1}{\|p(\lambda)\|_M} - \frac{1}{\Delta} = 0. \tag{234}$$

The following result describes some important properties of the function $\phi(\lambda)$ and provides the formula of a Newton iteration for (234).

Lemma 5.2.4 *Suppose $g \neq 0$. Then, $\phi(\lambda)$ is strictly increasing and concave on $(-\lambda_1, \infty)$. Moreover, the Newton iterate at λ with respect to (234) is*

$$\lambda^+ = \lambda + \left(\frac{\|p(\lambda)\|_M - \Delta}{\Delta}\right)\left(\frac{\|p(\lambda)\|_M^2}{p(\lambda)^T M H(\lambda)^{-1} M p(\lambda)}\right). \tag{235}$$

Proof. For $\lambda > -\lambda_1$, let $\bar{p}(\lambda) \equiv -\bar{H}(\lambda)^{-1}\bar{g}$, where $\bar{g} \equiv M^{-1/2}g$ and $\bar{H}(\lambda) \equiv M^{-1/2}H(\lambda)M^{-1/2} = M^{-1/2}HM^{-1/2} + \lambda I$. By (229) we have $\bar{p}(\lambda) = M^{1/2}p(\lambda)$, which together with (234) implies that

$$\phi(\lambda) = \frac{1}{\|\bar{p}(\lambda)\|} - \frac{1}{\Delta}.$$

Hence, by Lemma 7.3.1 on pp. 183 of [24], it follows that the function $\phi(\lambda)$ is strictly increasing and concave when $\lambda \in (-\lambda_1, \infty)$, and that its first derivative is given by

$$\phi'(\lambda) = \frac{\bar{p}(\lambda)^T \bar{H}(\lambda)^{-1} \bar{p}(\lambda)}{\|\bar{p}(\lambda)\|^3}. \tag{236}$$

The formula (235) for the Newton iteration $\lambda^+ = \lambda - \phi'(\lambda)/\phi(\lambda)$ can be easily derived using (236), $\bar{p}(\lambda) = M^{1/2} p(\lambda)$ and $\bar{H}(\lambda) = M^{-1/2} H(\lambda) M^{-1/2}$. ∎

The next result gives a few useful properties of Newton method applied to (234).

Proposition 5.2.5 *Suppose $g \neq 0$. Then the following statements hold:*

a) Suppose $\lambda \in (-\lambda_1, \lambda^*)$. Then all Newton iterates starting from λ will stay in $(-\lambda_1, \lambda^*)$ and converge to the solution λ^* of the equation (234) monotonically. The convergence is globally Q-linear with the ratio at least

$$\gamma_\lambda = 1 - \frac{\phi'(\lambda^*)}{\phi'(\lambda)} < 1$$

and is ultimately Q-quadratic.

b) Suppose $\lambda \in (\lambda^*, \infty)$. Then the next Newton iterate λ^+ satisfies

$$\lambda^+ \in (-\lambda_1, \lambda^*] \quad \text{or} \quad \lambda^+ \in (-\infty, -\lambda_1].$$

Proof. The proof is similar to the ones given in Lemmas 7.3.2 and 7.3.3 on pp. 185-186 of [24]. ∎

To compute the Newton iterate λ^+ according to (235), the NE method of [89] first computes the lower Cholesky factor L of $H(\lambda)$, and uses it to first compute a vector p such that $LL^T p = -g$ and then a vector w such that $Lw = Mp$. By (235), we then have

$$\lambda^+ = \lambda + \left(\frac{\|p\|_M - \Delta}{\Delta}\right)\left(\frac{\|p\|_M^2}{\|w\|^2}\right). \tag{237}$$

In our approach for solving the low rank version of (216) we entirely avoid the computation of Cholesky factorizations by instead computing the inverse of $H(\lambda)$ by means of the Sherman-Morrison-Woodbury (SMW) formula (see Subsection 3.2).

5.2.4 A safeguard Newton method

Since a Newton iteration might result in infeasible iterates $\lambda^+ \leq -\lambda_1$, the NE method of [89] uses some safeguard strategies to handle such iterates in order to obtain a globally convergent method for obtaining a NE solution of (216). The basic idea used is to bracket λ^* by a lower bound λ^L and an upper bound λ^U and reduce the length of the interval $[\lambda^L, \lambda^U]$ by using a clever bisection strategy. In this subsection, we discuss the details of this hybrid method.

At every iteration of the method, we have two scalars λ^L and λ^U such that $0 \leq \lambda^L \leq \lambda^* \leq \lambda^U$ and a current approximation $\lambda \in [\lambda^L, \lambda^U]$ of λ^*. Each iteration of the method then consists of updating the quantities λ^L, λ^U and λ. We first describe the basic idea used to update λ^L. Suppose that $u \in \Re^n$ is a vector such that $\|u\|_M = 1$. Using (228), we see that for any $\lambda \in \Re$,

$$u^T H(\lambda) u = (M^{1/2}u)^T (M^{-1/2} H M^{-1/2} + \lambda I)(M^{1/2}u) \geq (\lambda_1 + \lambda)\|u\|_M^2 = \lambda_1 + \lambda.$$

Defining

$$\lambda^B = \lambda^B(\lambda, u) \equiv \lambda - u^T H(\lambda) u, \tag{238}$$

it follows from the above inequality that $\lambda^B \leq -\lambda_1 \leq \lambda^*$, or in words, λ_B is a lower bound for λ^*. In view of this discussion, we conclude that a natural update for λ^L is simply to let $\lambda^L \leftarrow \max(\lambda^L, \lambda^B)$.

We are now ready to describe how the NE method of [89] updates the three quantities λ^L, λ^U and λ. Fix some constant $\theta \in (0, 1)$ (e.g., $\theta = 0.01$). It is convenient to consider the following three cases separately:

i) Assume $\lambda \leq -\lambda_1$. If $\lambda = -\lambda_1$, we perform the update $\lambda^L \leftarrow \max(\lambda^L, \lambda)$. Otherwise, $H(\lambda)$ is indefinite, and hence there exists a vector u such that $\|u\|_M = 1$ and $u^T H(\lambda) u < 0$. One approach to find such a vector u is to perform a partial Cholesky factorization of $H(\lambda)$ to find a scalar $\delta > 0$ and a vector v such that

$$(H(\lambda) + \delta e_k e_k^T)v = 0 \text{ and } e_k^T v = 1. \tag{239}$$

(for more details, see [89] or pp. 191-192 of [24]). Letting $u = v/\|v\|_M$ in (238) and using (239), we easily see that $\lambda^B = \lambda + \delta/\|v\|_M^2$. Then, we perform the update $\lambda^L \leftarrow \max(\lambda^L, \lambda^B)$. Finally, we perform the update $\lambda \leftarrow \max(\sqrt{\lambda^L \lambda^U}, \lambda^L + \theta(\lambda^U - \lambda^L))$.

ii) Assume $\lambda \in (-\lambda_1, \lambda^*)$. In this case, we perform the updates $\lambda^L \leftarrow \lambda$ and $\lambda \leftarrow \lambda^+$, where λ^+ denotes the Newton iterate defined in (235).

iii) Assume $\lambda > \lambda^*$. In this case, we have seen in the paragraph following Lemma 5.2.3 that a vector $u \in \Re^n$ such that $\|u\|_M = 1$ can be computed using the LINPACK technique which makes $u^T H(\lambda) u$ small as long as $\lambda + \lambda_1$ is small. This vector u, together with λ, is then used to compute λ^B according to (238), and the updates $\lambda^U \leftarrow \lambda$, $\lambda^L \leftarrow \max(\lambda^L, \lambda^B)$ and $\lambda \leftarrow \max(\lambda^+, \lambda^L)$ are then performed.

The above scheme for updating λ^L, λ^U and λ can be shown to generate a sequence of λ's which approaches λ^*. Indeed, if case ii) occurs then the sequence of λ's generated afterwards approaches λ^* monotonically from the left in view of Proposition 5.2.5(a). Also, Proposition 5.2.5(b) implies that if case iii) occurs then either case i) or ii) must occur at the next iteration. Hence, if case ii) never occurs then case i) must occur infinitely often. But every time i) occurs, it is easy to see that the ratio of the length of the interval $[\lambda^L, \lambda^U]$ at the end of the next iteration and its length at the current iteration is bounded above by $\max(\theta, 1-\theta)$. Hence, if case ii) never occurs, the length of the generated intervals $[\lambda^L, \lambda^U]$ converges to zero, and thus the generated sequence of λ's approaches λ^*.

Note that the implementations of cases i) and iii) of the above scheme for updating λ^L, λ^U and λ are based on the computation of the (partial) Cholesky factorization of a matrix. In our approach for finding NE solutions of the low rank version of (216), these cases must be implemented differently so as to avoid the computation of Cholesky factorizations, which are known to be expensive for large scale problems. These alternative implementations of cases i) and iii) are discussed in detail in Subsections 5.3.2 and 5.3.3.

We are now ready to state the whole algorithm of [89] for finding a NE solution of (216).

Algorithm 1 (NE method for solving (216)):

Let constants θ, k_e, $k_h \in (0,1)$ be given (e.g., $\theta = 0.01$, $k_e = 0.1$, $k_h = 0.2$).

1) Find initial scalars $0 \leq \lambda^L < \lambda^U$ such that $\lambda^* \in [\lambda^L, \lambda^U]$ and set $\lambda = \lambda^L$.

2) Attempt to do Cholesky factorization $H(\lambda) = LL^T$ to check whether $\lambda > -\lambda_1$.

3) If $\lambda > -\lambda_1$, solve $LL^T p = -g$ for p. Check for interior convergence and easy termination.

If $\lambda > \lambda^*$, compute a pair $(u, \alpha) \in \Re^n \times \Re$ to check for the hard termination (231).

4) If $\lambda \leq -\lambda_1$, then update λ^L and λ according to case i) as above, and go to step 2).

5) If $\lambda \in (-\lambda_1, \lambda^*)$, then update λ^L and λ according to case ii) as above, and go to step 2).

6) If $\lambda > \lambda^*$, then update λ^L, λ^U and λ according to case iii) as above, and go to step 2).

End

For large-scale problems, several computational difficulties arise in the NE method of [89] above. In step 1), it generally takes $\mathcal{O}(n^2)$ amount of arithmetic operations to find the initial λ^U (see Section 7.3.8 of [24]), which is somehow expensive for large-scale problems. The Cholesky or partial Cholesky factorization of $H(\lambda)$ used in steps 2), 4), 6) and 7) needs $\mathcal{O}(n^3)$ amount of arithmetic operations. It will be prohibitive for large-scale problems. In the next section, we will modify the NE method of [89] above to overcome those difficulties for solving the LRTR subproblem.

5.3 A modified NE method for solving LRTR subproblem

In this section a modified NE (MNE) method for solving large-scale LRTR subproblems is presented. We follow the framework of Algorithm 1 as described in Section 5.2. Our main effort is to overcome the computational difficulties mentioned in Section 5.2.

This section is divided into four subsections. A more efficient approach for checking whether $H(\lambda)$ is positive definite is given in Subsection 5.3.1. We also modify the approach of solving the linear equation $H(\lambda)p = -g$ and computing a Newton iterate in this subsection. A more efficient approach for dealing with hard case termination is developed in Subsection 5.3.2. In Subsection 5.3.3, a more efficient approach for improvingλ^L when $\lambda < \lambda_1$ is given. In Subsection 5.3.4, we develop a cheaper approach to initialize λ^U. We

emphasize that all modified approaches completely avoid computing Cholesky or partial Cholesky factorization of large-scale matrices.

5.3.1 Checking positive definiteness of $H(\lambda)$ and solving $H(\lambda)p = -g$

In step 2) of Algorithm 1, given any $\lambda \geq 0$, the NE method of [89] checks whether $H(\lambda)$ is positive definite by computing the Cholesky factorization of it, which is very expensive and even prohibitive for the large-scale problems. In this subsection, we provide a more efficient method instead, which needs $\mathcal{O}(n)$ amount of arithmetic operation for large-scale LRTR problems. Furthermore, in steps 5) and 6) of Algorithm 1, the NE method of [89] uses the Cholesky factor of $H(\lambda)$ to solve the linear equation $H(\lambda)p = -g$ and compute Newton iterate, respectively. We will modify those approaches as well in this subsection.

The following theorem provides the main tool for the analysis in this subsection.

Theorem 5.3.1 *Let $\hat{E} \in \mathcal{S}^m$, $\hat{V} \in \Re^{n\times m}$ and an invertible matrix $\hat{D} \in \mathcal{S}^n$ be given and define $\hat{H} \equiv \hat{D} + \hat{V}\hat{E}\hat{V}^T$ and $\hat{W} \equiv \hat{V}^T\hat{D}^{-1}\hat{V}$. Then, the following statements hold:*

i) If $\hat{E}$ is invertible, then $\hat{H}$ is invertible if and only if $\hat{E}^{-1} + W$ is invertible.

ii) If $\hat{D} \succ 0$, then $\hat{H} \succeq 0$ if and only if $\hat{W} + \hat{W}\hat{E}\hat{W} \succeq 0$.

iii) If $\hat{D} \succ 0$, then $\hat{H} \succ 0$ and $\hat{V}$ has full column rank if and only if $\hat{W} + \hat{W}\hat{E}\hat{W} \succ 0$.

Proof. It is well-known that if $\hat{E}^{-1} + \hat{W}$ is invertible then the SWM formula applied to $\hat{H}$ implies that $\hat{H}^{-1} = \hat{D}^{-1} - \hat{D}^{-1}\hat{V}(\hat{E}^{-1} + \hat{W})^{-1}\hat{V}^T\hat{D}^{-1}$. Similarly, if $\hat{H} = \hat{D} + \hat{V}\hat{E}\hat{V}^T$ is invertible then the SMW formula applied to $\hat{E}^{-1} + \hat{W}$ reveals that this matrix is invertible. Hence, i) follows.

To prove statement ii), assume that $\hat{D} \succ 0$. We can then write $\hat{H}$ as $\hat{H} = \hat{D}^{1/2}F\hat{D}^{1/2}$, where $F \equiv I + \hat{D}^{-1/2}\hat{V}\hat{E}\hat{V}^T\hat{D}^{-1/2}$. Clearly, $\hat{H} \succeq 0$ if and only if $F \succeq 0$. In view of the decomposition $\Re^n = \mathrm{Ker}(\hat{V}^T\hat{D}^{-1/2}) + \mathrm{Im}(\hat{D}^{-1/2}\hat{V})$, it follows that, for any $p \in \Re^n$, there exist u, $v \in \Re^n$ and $y \in \Re^m$ such that

$$p = u + v, \quad \hat{V}^T\hat{D}^{-1/2}u = 0, \quad v = \hat{D}^{-1/2}\hat{V}y.$$

This relation, the facts that $Fu = u$ and $u^T v = 0$, and some simple algebraic manipulation imply that

$$\begin{aligned} p^T F p &= (u+v)^T F (u+v) = u^T F u + 2u^T F v + v^T F v, \\ &= \|u\|^2 + v^T F v = \|u\|^2 + y^T (\hat{W} + \hat{W}\hat{E}\hat{W}) y. \end{aligned}$$

By the arbitrariness of p, we easily see that $F \succeq 0$ if only if $\hat{W} + \hat{W}\hat{E}\hat{W} \succeq 0$. This, together with the fact that $\hat{H} \succeq 0$ if only if $F \succeq 0$, implies that ii) holds. Under additional assumption that $\hat{V}$ has full column rank, the statement iii) can be shown by using a similar argument as ii). ∎

We now describe an efficient approach to check whether $H(\lambda)$ with $\lambda \geq 0$ is positive definite in step 2) of Algorithm 1. Noting that $H(\lambda) = H + \lambda M$, we see from (218) and (219) that

$$H(\lambda) = \hat{D} + \hat{V}\hat{E}\hat{V}^T, \tag{240}$$

where

$$\hat{D} \equiv D + \lambda \tilde{D} \succ 0, \quad \hat{V} \equiv (V, \tilde{V}), \quad \hat{E} \equiv \mathrm{Diag}(E, \lambda \tilde{E}). \tag{241}$$

Hence, by Theorem 5.3.1 ii), if $\hat{V}$ has full column rank, then we can check the positive definiteness of $H(\lambda)$ by checking whether

$$X \equiv \hat{W} + \hat{W}\hat{E}\hat{W} \tag{242}$$

is positive definite. On the other hand, if $\hat{V}$ does not have full column rank, then we determine a matrix R with full column rank and a matrix T such that $\hat{V} = RT$. It then follows that $H(\lambda) = \hat{D} + \hat{V}\hat{E}\hat{V}^T = \hat{D} + R\check{E}R^T$, where $\check{E} \equiv T\hat{E}T^T$. Hence, Theorem 5.3.1 ii) can now be used to check the positive definiteness of $H(\lambda)$ by checking whether $\check{W} + \check{W}\check{E}\check{W} \succ 0$, where $\check{W} \equiv R^T \hat{D}^{-1} R$. For convenience of the presentation, we will assume throughout the remaining subsections that $\hat{V}$ has full column rank.

Recall that one of the requirements for (216) to be a LRTR subproblem is that the number of columns of $\hat{V}$ be small. In this case, X is a small-sized matrix whose positive definiteness can be checked by performing a relatively cheap Cholesky factorization. Since

the amount of arithmetic operations to compute X for a large-scale LRTR subproblem is $\mathcal{O}(n)$, the above approach for checking whether $H(\lambda)$ is positive definite only requires $\mathcal{O}(n)$ arithmetic operations. If $H(\lambda)$ turns out to be positive definite, we can then solve the linear system $H(\lambda)p = -g$ by means of SMW formula as

$$\begin{aligned} p &= -H(\lambda)^{-1}g = -\left(\hat{D}^{-1} - \hat{D}^{-1}\hat{V}(\hat{E}^{-1} + \hat{V}^T\hat{D}^{-1}\hat{V})^{-1}\hat{V}^T\hat{D}^{-1}\right)g, \\ &= -\hat{D}^{-1}g + \hat{D}^{-1}\hat{V}(\hat{E}^{-1} + \hat{W})^{-1}\hat{V}^T\hat{D}^{-1}g. \end{aligned}$$

Note that the matrix $\hat{E}^{-1} + \hat{W}$ is invertible due to the fact $H(\lambda) \succ 0$ and Theorem 5.3.1 i). Since the size of $\hat{E}^{-1} + \hat{W}$ is small, this approach for solving the linear system $H(\lambda)p = -g$ also requires $\mathcal{O}(n)$ arithmetic operations. Note that, in the context of a LRTR subproblem, the Newton iterate λ^+ can be efficiently computed by means of (235), which requires solving another linear system with coefficient matrix $H(\lambda)$.

5.3.2 Handling the hard case termination

Recall that one of the key parts in the implementation of steps 3) and 6) of Algorithm 1 is the computation of a vector $u = u_\lambda$ such that $\|u\|_M = 1$ and $u^T H(\lambda)u$ approaches zero as $\lambda \downarrow -\lambda_1$ (see Subsections 5.2.2 and 5.2.4). In this subsection, we provide an efficient approach to find such a vector u in the context of the low-rank version of (216), which completely avoids the computation of the Cholesky factorization of $H(\lambda)$.

Recall from (218) that $H = D + VEV^T$, where $D \succ 0$ and E is diagonal and nonsingular. We can partition E (after performing a symmetric permutation of its rows and columns) as $E = \mathrm{Diag}(E_1, -E_2)$, where both E_1 and E_2 are positive diagonal matrices. Accordingly, we partition V as $V = (V_1, V_2)$, and hence

$$VEV^T = V_1E_1V_1^T - V_2E_2V_2^T. \tag{243}$$

Noting that $H(\lambda) = H + \lambda M$, we can write $H(\lambda)$ as

$$H(\lambda) = F(\lambda) - V_2E_2V_2^T, \tag{244}$$

where $F(\lambda) = D + \lambda M + V_1E_1V_1^T \succ 0$ for any $\lambda \geq 0$ due to the fact that D, $M \succ 0$.

The following technical lemma provides the key tool for our analysis in this subsection.

Lemma 5.3.2 *Assume that λ_1 defined in (228) is nonpositive. Then,*

$$\lim_{\lambda \downarrow -\lambda_1} \lambda_{\max}\left(V_2^T H(\lambda)^{-1} V_2\right) = \infty,$$

where V_2 is defined as above.

Proof. For any $\lambda > -\lambda_1$, using (244) and SMW formula twice, we have

$$H(\lambda)^{-1} = F(\lambda)^{-1} + F(\lambda)^{-1} V_2 \Big(E_2^{-1} - {V_2}^T F(\lambda)^{-1} V_2\Big)^{-1} {V_2}^T F(\lambda)^{-1}, \tag{245}$$

and

$$\begin{aligned} \Big(E_2^{-1} - {V_2}^T F(\lambda)^{-1} V_2\Big)^{-1} &= E_2 + E_2 {V_2}^T \Big(F(\lambda) - V_2 E_2 {V_2}^T\Big)^{-1} V_2 E_2, \\ &= E_2 + E_2 {V_2}^T H(\lambda)^{-1} V_2 E_2. \end{aligned} \tag{246}$$

Using the definition of $F(\lambda)$ and the fact $M \succ 0$, we easily see that $F(\lambda) \succ D \succ 0$ for any $\lambda > -\lambda_1 \geq 0$. This implies that $\|F(\lambda)^{-1}\|$, and hence $\|F(\lambda)^{-1} V_2\|$, is bounded for all $\lambda > -\lambda_1$. From (245), we obtain that

$$\|H(\lambda)^{-1}\| \leq \|F(\lambda)^{-1}\| + \|F(\lambda)^{-1} V_2\| \ \left\|\Big(E_2^{-1} - {V_2}^T F(\lambda)^{-1} V_2\Big)^{-1}\right\| \ \|{V_2}^T F(\lambda)^{-1}\| \tag{247}$$

By the definitions of $H(\lambda)$ and λ_1, we have $\lim_{\lambda \downarrow -\lambda_1} \|H(\lambda)^{-1}\| = \infty$, which, together with (247) and the fact that $\|F(\lambda)^{-1}\|$ and $\|F(\lambda)^{-1} V_2\|$ are bounded for all $\lambda > -\lambda_1 \geq 0$, implies

$$\lim_{\lambda \downarrow -\lambda_1} \left\|\Big(E_2^{-1} - {V_2}^T F(\lambda)^{-1} V_2\Big)^{-1}\right\| = \infty. \tag{248}$$

Moreover, from (246), we have

$$\left\|\Big(E_2^{-1} - {V_2}^T F(\lambda)^{-1} V_2\Big)^{-1}\right\| \leq \|E_2\| + \|E_2\| \ \left\|{V_2}^T H(\lambda)^{-1} V_2\right\| \ \|E_2\|,$$

with together with (248) implies that

$$\lambda_{\max}\left(V_2^T H(\lambda)^{-1} V_2\right) = \left\|{V_2}^T H(\lambda)^{-1} V_2\right\| \to \infty \text{ as } \lambda \downarrow -\lambda_1.$$

■

The following theorem provides an efficient approach to compute the vector u for dealing with the hard case termination in step 3) of Algorithm 1 and updating λ^L in step 6) of Algorithm 1.

Theorem 5.3.3 *Assume that λ_1 defined in (228) is nonpositive. Suppose that*

$$u_\lambda = H(\lambda)^{-1}v/\|H(\lambda)^{-1}v\|_M,$$

where $v = V_2 r$ and r is a unit eigenvector of $V_2^T H(\lambda)^{-1} V_2$ corresponding to its maximum eigenvalue. Then,

$$\lim_{\lambda \downarrow -\lambda_1} u_\lambda^T H(\lambda) u_\lambda = 0.$$

Proof. It follows from Cauchy-Schwarz inequality that

$$v^T H(\lambda)^{-1} v \leq \|v\|_{M^{-1}} \|H(\lambda)^{-1} v\|_M. \tag{249}$$

Using (249) and the definitions of v and r, we have

$$\begin{aligned} u_\lambda^T H(\lambda) u_\lambda &= \frac{v^T H(\lambda)^{-1} v}{\|H(\lambda)^{-1} v\|_M^2} = \frac{(v^T H(\lambda)^{-1} v)^2}{(v^T H(\lambda)^{-1} v)\|H(\lambda)^{-1} v\|_M^2}, \\ &\leq \frac{\|v\|_{M^{-1}}^2}{v^T H(\lambda)^{-1} v} = \frac{r^T V_2^T M^{-1} V_2 r}{r^T V_2^T H(\lambda)^{-1} V_2 r}, \\ &\leq \frac{\|V_2^T M^{-1} V_2\|}{\lambda_{\max}(V_2^T H(\lambda)^{-1} V_2)}, \end{aligned}$$

which, together with Lemma 5.3.2, immediately implies that the conclusion holds. ∎

Before ending this subsection, we make two observations. First, since $\lambda^* = -\lambda_1$ in the hard case, Theorem 5.3.3 implies that the vector u_λ defined in its statement satisfies $\lim_{\lambda \downarrow \lambda^*} u_\lambda^T H(\lambda) u_\lambda = 0$, which is exactly the condition required in the discussion of the hard case (see Subsection 5.2.2). Second, since the number of columns of V is assumed to be small in the low-rank version subproblem (216), it follows that the matrix $V_2^T H(\lambda)^{-1} V_2$ is small-sized, and hence a unit eigenvector r as in Theorem 5.3.3 can be easily computed. Moreover, the SMW formula can be used to compute $V_2^T H(\lambda)^{-1} V_2$ and u_λ in $\mathcal{O}(n)$ arithmetic operations.

5.3.3 Improving λ^L when $\lambda < -\lambda_1$

Recall that the key part in the implementation of step 4) of Algorithm 1 consists of finding a vector u satisfying $\|u\|_M = 1$ and $u^T H(\lambda) u < 0$, whenever $\lambda < -\lambda_1$ (see Subsections 5.2.4). In this subsection, we provide an efficient approach to find such a vector u in the

context of the low-rank version of (216), which completely avoids the computation of a partial Cholesky factorization of $H(\lambda)$.

Assume then that $0 \leq \lambda < -\lambda_1$. This implies that $H(\lambda)$ is indefinite, and hence that the matrix X defined in (242) is also indefinite, in view of Theorem 5.3.1(iii). Hence, letting y be an eigenvector of X corresponding to its minimum eigenvalue, we have that $y^T X y < 0$. Using the definition of $\hat{W}$ and relations (240), (241) and (242), we easily see that

$$X = \hat{W} + \hat{W}\hat{E}\hat{W} = \hat{V}^T \hat{D}^{-1} H(\lambda) \hat{D}^{-1} \hat{V}. \tag{250}$$

Hence, letting $u := \hat{D}^{-1}\hat{V}y / \|\hat{D}^{-1}\hat{V}y\|_M$, we have that $\|u\|_M = 1$ and

$$u^T H(\lambda) u = \frac{y^T X y}{\|\hat{D}^{-1}\hat{V}y\|_M^2} < 0.$$

Note that since X is a small-sized matrix (see Subsection 5.3.1), it is relatively cheap to compute the vector y as described above. Moreover, since the amount of arithmetic operations to compute X for a large-scale LRTR subproblem is $\mathcal{O}(n)$, the computation of the vector u described above can be carried out in $\mathcal{O}(n)$ arithmetic operations.

5.3.4 Finding the initial λ^U

Recall that step 1) of Algorithm 1 requires initial estimates of the lower bound λ^L and the upper bound λ^U. An approach for estimating these bounds for a general TR subproblem in $\mathcal{O}(n^2)$ arithmetic operations is described in Section 7.3.8 of [24]. For the large-scale lower-rank version subproblem (216), the above approach is expensive, and hence not suitable.

In our implementation of Algorithm 1, we set $\lambda^L = 0$. We now provide an efficient approach to find an initial estimate of λ^U in the context of the low-rank version of (216) in this subsection.

Recall that H and M have low-rank structure (see (218) and (219)). Assume that the row size of matrices E and $\tilde{E}$ is $\bar{k}$ and $\tilde{k}$, respectively. For the convenience of the presentation, we rewrite H and M in (218) and (219) as follows

$$H = D + \sum_{i=1}^{\bar{k}} E_{ii} v_i v_i^T, \tag{251}$$

$$M = \tilde{D} + \sum_{i=1}^{\tilde{k}} \tilde{E}_{ii} \tilde{v}_i \tilde{v}_i^T, \tag{252}$$

where v_i, $\tilde{v}_i$ are the ith column of V and $\tilde{V}$, respectively.

Given any $\epsilon > 0$, we can trivially set initial λ^U to be ϵ if $\lambda^* = 0$. Hence, we now assume that $\lambda^* > 0$. It follows from Lemma 5.2.1 that λ^* together with a global solution p of (216) satisfies

$$\begin{aligned} (H + \lambda^* M)p &= -g, & (253)\\ \|p\|_M &= \Delta, & (254)\\ H + \lambda^* M &\succeq 0 \quad . & (255) \end{aligned}$$

Multiplying (253) by p^T on the left and using (254) and the fact $M \succ 0$, we obtain

$$p^T H p + \lambda^* \Delta^2 = -p^T g \leq \|p\|_M \|g\|_{M^{-1}} = \Delta \|g\|_{M^{-1}}.$$

Hence

$$\lambda^* \leq \|g\|_{M^{-1}} \Delta^{-1} - (p^T H p)\Delta^{-2}. \tag{256}$$

Let $\tilde{p} = M^{1/2} p/\Delta$. Noting that $\|\tilde{p}\| = 1$, we have

$$\begin{aligned} (p^T H p)\Delta^{-2} &= \tilde{p}^T M^{-1/2} H M^{-1/2} \tilde{p},\\ &\geq \tilde{p}^T M^{-1/2} \left(D + \sum_{\{i | E_{ii} < 0\}} E_{ii} v_i v_i^T \right) M^{-1/2} \tilde{p},\\ &\geq \lambda_{\min}(D) \|M^{-1/2}\tilde{p}\|^2 + \sum_{\{i | E_{ii} < 0\}} E_{ii} (v_i^T M^{-1/2} \tilde{p})^2,\\ &\geq \zeta \|M^{-1/2}\tilde{p}\|^2, \end{aligned} \tag{257}$$

where $\zeta = \lambda_{\min}(D) + \sum_{\{i | E_{ii} < 0\}} E_{ii} \|v_i\|^2$. Using (252) and the fact that $\tilde{E}_{ii} > 0$ for all i, we have

$$\lambda_{\min}(M) \geq \min_{1 \leq i \leq n} \tilde{D}_{ii} \ \text{ and } \ \lambda_{\max}(M) \leq \max_{1 \leq i \leq n} \tilde{D}_{ii} + \sum_{i=1}^{\tilde{k}} \tilde{E}_{ii} \|\tilde{v}_i\|^2. \tag{258}$$

Note that

$$\lambda_{\max}(M)^{-1} \leq \|M^{-1/2}\tilde{p}\|^2 \leq \lambda_{\min}(M)^{-1}.$$

This together with (258) implies that

$$\left(\max_{1 \leq i \leq n} \tilde{D}_{ii} + \sum_{i=1}^{\tilde{k}} \tilde{E}_{ii} \|\tilde{v}_i\|^2 \right)^{-1} \leq \|M^{-1/2}\tilde{p}\|^2 \leq \left(\min_{1 \leq i \leq n} \tilde{D}_{ii} \right)^{-1}. \tag{259}$$

Using (256), (257), and (259), we see that $\lambda^* \leq \lambda^e$, where λ^e is defined as

$$\lambda^e \equiv \begin{cases} \|g\|_{M^{-1}}\Delta^{-1} - \zeta\left(\max\limits_{1\leq i\leq n} \tilde{D}_{ii} + \sum\limits_{i=1}^{\tilde{k}} \tilde{E}_{ii}\|\tilde{v}_i\|^2\right)^{-1} & \text{if } \zeta \geq 0, \\ \\ \|g\|_{M^{-1}}\Delta^{-1} - \zeta\left(\min\limits_{1\leq i\leq n} \tilde{D}_{ii}\right)^{-1} & \text{if } \zeta < 0. \end{cases} \tag{260}$$

This together with the fact that $\lambda^U = \epsilon$ if $\lambda^* = 0$ implies that $\max(\epsilon, \lambda^e)$ is a proper initial estimate of the upper bound λ^U for any λ^*.

Using the fact that M (219) has low-rank structure, we see that $M^{-1}g$ can be computed by means of SMW formula requiring $\mathcal{O}(n)$ arithmetic operations. This together with (260), and the fact that $\|g\|_{M^{-1}} = \sqrt{g^T M^{-1} g}$ and $\bar{k}$ and $\tilde{k}$ are small, implies that this approach for finding initial estimate of the upper bound λ^U requires $\mathcal{O}(n)$ arithmetic operations in the context of a LRTR subproblem.

5.4 Some numerical implementation results

In this section, our main goal is to test the numerical performance of "low-rank" trust region methods whose LRTR subproblems are solved by the MNE method proposed in Section 5.3. For this purpose, we implement a specific version of the modified log-barrier (MLB) algorithm due to Polyak [101] (see Subsection 5.4.1) and use it to solve a collection of nonlinear programming problems from CUTEr [43] where only simple bound constraints are present. Like the log-barrier method discussed in Section 5.1, the MLB algorithm also consists of solving a parametrized family of unconstrained nonlinear problems. In our implementation, these subproblems are solved by using a "low-rank" trust region approach similar to the one discussed in Section 5.1 in the context of the log-barrier method. This section is divided into two subsections. In Subsection 5.4.1, we discuss the generic MLB algorithm for solving nonlinear programming problems with general inequality constraints and its specialization to problems with simple bound constraints. In Subsection 5.4.2, we report the computational results of our implementation of the MLB method and its comparison with a version of LANCELOT [23] based on the forementioned collection of problems from CUTEr [43].

5.4.1 The modified log-barrier algorithm

In this subsection, we discuss the generic MLB algorithm for solving nonlinear programming problems with general inequality constraints and its specialization to problems with simple bound constraints.

Consider the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \geq 0,\ i = 1, \cdots, m. \end{array} \tag{261}$$

where the functions $f(x)$ and $c_i(x)$, $i = 1, \cdots, m$ are twice continuously differentiable in $\Re^n$. Its associated first-order optimality conditions are:

$$\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0,\ \lambda \geq 0,$$
$$\sum_{i=1}^{m} \lambda_i c_i(x) = 0,\ c_i(x) \geq 0,\ i = 1, \cdots, m.$$

The MLB method proposed by Polyak [101] consists of solving a sequence of unconstrained problems with objective functions given by

$$\mathcal{M}(x, \mu^{(k)}, \lambda^{(k)}) = f(x) - \mu^{(k)} \sum_{i=1}^{m} \lambda_i^{(k)} \log\left(\frac{c_i(x)}{\mu^{(k)}} + 1\right), \tag{262}$$

where $\lambda^{(k)} \in \Re^m$ is an estimate of a Lagrange multiplier at a solution of (261) and $\mu^{(k)} > 0$ is a log-barrier parameter. Letting $x^{(k)}$ denote a stationary point of $\mathcal{M}(x, \mu^{(k)}, \lambda^{(k)})$, Polyak [101] has shown under reasonable conditions that there exists a threshold value $\bar{\mu} > 0$ such that for any fixed $\mu \in (0, \bar{\mu})$, the MLB method which updates $\{\lambda^{(k)}\}$ according to

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} / (c_i(x^{(k)})/\mu + 1), \tag{263}$$

generates a sequence of iterates $\{(x^{(k)}, \lambda^{(k)})\}$ which converges to a point satisfying the first-order optimality condition of (261) as $k \to \infty$.

In order to give the detailed description of the MLB method, we introduce the following parameters and definitions (see Breitfeld and Shanno [16] and [17]).

Let $T_1 = 10^{-4}$, $T_2 = 10^{-6}$, $\epsilon_0 = 10^{-5}$, $\sigma = 0.5$ or 0.1. Also, let

$$r \ = \ -0.5\log_{10}(T_1), \quad \epsilon_k = \max(\epsilon_0, 10^{-(k+r-1)}),$$

$$\begin{aligned}
\nu_1^{(k)} &= \max\left\{\frac{\|\nabla \mathcal{M}\|}{1+\|x^{(k)}\|}, -\min_{i=1,\cdots,m} c_i(x^{(k)})\right\}, \\
\nu_2^{(k)} &= \max\left\{\nu_1^{(k)}, \frac{\sum_{i=1}^m \lambda_i^{(k+1)}|c_i(x^{(k)})|}{1+\|x^{(k)}\|}, \frac{\|\nabla f(x^{(k)}) - \sum_{i=1}^m \lambda_i^{(k+1)} \nabla c_i(x^{(k)})\|_\infty}{1+\|x^{(k)}\|}\right\} \quad (264) \\
\nu_3^{(k)} &= \frac{|f(x^{(k)}) - f(x^{(k-1)})|}{1+|f(x^{(k-1)})|}, \quad \nu_4^{(k)} = \frac{|\nu_2^{(k)} - \nu_2^{(k-1)}|}{\nu_2^{(k-1)}}.
\end{aligned}$$

The algorithm below is a complete description of MLB method as implemented in [16] and [17].

Algorithm MLB:

Let $\xi \in (0,1)$, $\mu^{(1)} > 0$, $\lambda^{(1)} > 0$, $\nu_1^{(0)}$, $\nu_2^{(0)}$, $\nu_3^{(0)}$ be given, and set $k = 1$.

For $k = 1, 2, 3, \ldots$, until $\nu_2^{(k-1)} < T_1$ or ($\nu_1^{(k-1)} < T_1$ and $\nu_3^{(k-1)} < T_2$), **do**

1. Minimize the function (262) approximately,

   obtaining $x^{(k)}$ such that $\|\nabla \mathcal{M}(x^{(k)}, \mu^{(k)}, \lambda^{(k)})\| \leq \epsilon_k$.

2. Update $\lambda^{(k+1)}$ by (263). If $\nu_4^{(k)} < \xi$, set $\mu^{(k+1)} = \sigma \mu^{(k)}$.

   Increment k by 1, and return to step 1.

End

In our implementation, we initialize $\mu^{(0)} = 10^{-2}$ and $\lambda_i^{(0)} = 1$ for $i = 1, \cdots, m$ as suggested in [16] and [101], respectively.

While Breitfeld and Shanno [16] have used a line search method to find $x^{(k)}$, we instead use a trust region method whose associated TR subproblems have quadratic objective functions with low-rank Hessian matrices obtained by means of limited-memory BFGS (L-BFGS) method (see Section 9.1 of [98])). The MNE method is then used to solve the resulting LRTR subproblems.

We next provide more details of how a low-rank approximation of the Hessian of $\mathcal{M}(\cdot, \mu^{(k)}, \lambda^{(k)})$ is computed in the context of solving nonlinear programming problems with simple bound constraints. For the purpose of this discussion, assume that the constraints of (261) are given by $c_i^l(x) \equiv x_i - l_i \geq 0$ for $i \in I_l$ and $c_i^u(x) \equiv u_i - x_i \geq 0$ for $i \in I_u$ where $I_l, I_u \subseteq \{1, \cdots, n\}$ are index sets corresponding to the lower and upper bound constraints, respectively. Let λ^l and λ^u be the Lagrange multipliers corresponding to the lower and upper

bound constraints, respectively. We easily see from (262) that $\nabla^2\mathcal{M}(x,\mu,\lambda) = \nabla^2 f(x)+\mu Q$, where

$$Q = \sum_{i\in I_l} \frac{\lambda_i^l e_i e_i^T}{(c_i^l(x)+\mu)^2} + \sum_{i\in I_u} \frac{\lambda_i^u e_i e_i^T}{(c_i^u(x)+\mu)^2}. \tag{265}$$

Since Q is a diagonal matrix and can be easily computed, it makes sense to just compute a low-rank approximation F of $\nabla^2 f(x)$ and use $H \equiv F + \mu Q$ as a low-rank approximation of $\nabla^2\mathcal{M}(x,\mu,\lambda)$. We use the L-BFGS method to obtain a low-rank apprximation of $\nabla^2 f(x)$. This matrix H is used as the Hessian of the objective function of the LRTR subproblem (216) at the point x. Moreover, the other data of this subproblem is determined as $g \equiv \nabla\mathcal{M}(x,\mu,\lambda)$ and $M \equiv \mathrm{Diag}(v)$, where $v \in \Re^n$ is defined as $v_i = 1/\sqrt{Q_{ii}}$ if $i \in I_l \cup I_u$, and $v_i = 1$ if $i \notin I_l \cup I_u$.

5.4.2 Implementations of some problems from CUTEr

In this subsection, we will report the computational results obtained from the implementation of a specifc version of Algorithm MLB described in Subsection 5.4.1 and present the comparisons of our method with LANCELOT [23] on a collection of nonlinear programming problems from CUTEr [43] where only simple bound constraints are present.

All computations are performed on a Sun Ultra 10 workstation which has a single UltraSPARC IIi processor running at 440Mhz and 512MB of memory. The sixty test problems are selected from CUTEr. Seventeen of them have simple bound constraints and fixed variables. The remaining problems are unconstrained ones. Each row of Tables 4 and 5 gives the problem name, the number of variables, the number of bound constraints, the number of free variables, and the number of fixed variables on columns one through five, respectively.

Some computational results for our code are also presented in Tables 4 and 5. For each test problem, the total number of iterations performed by the MNE method is given in the sixth column, and the total number of LRTR subproblems (216) solved is given in the seventh column. The average iterations performed by the MNE method for each test problem is given in the eighth column, which is obtained by dividing the entry in the sixth column by the entry in the seventh clolumn.

From the eighth column of Tables 4 and 5, we observe that the average iterations

performed by the MNE method are between 1.0 and 2.0 for almost all test problems, except the problems QRTQUAD (2.89) and FLETCBV2 (0). Moreover, the average of the entries in the eighth column over all test problems is 1.25, which is better than the average 1.6 obtained by Moré and Sorensen [89] for the NE method applied to the general TR subproblems (216). Hence, this indicates that the MNE method developed in this chapter is efficient and robust for solving the low-rank version TR subproblem (216).

Our code MTR is written in ANSI C and LANCELOT is a FORTAN code. MTR and LANCELOT are both compiled under the default optimization. In MTR, we set the parameter $\sigma = 0.1$ for the Algorithm MLB and store 3 most recent vector pairs that provide curvature information for L-BFGS update. We select the same initial point x^0 as LANCELOT whenever it is strictly feasible; otherwise, we modify the infeasible components of this initial point x^0 in order to make it strictly feasible. We set up an upper bound of one hour computation time (or 3,600 seconds) per problem for both codes. We implement the version of LANCELOT which uses bfgs-approximate-second-derivatives, bandsolver-preconditioned-cg-solver, inexact-cauchy-point, two-norm-trust-region and all its other default settings. LANCELOT terminates when the infinity norm of the projected gradient is less than 10^{-5} or exceeds one hour computation time.

Tables 6 and 7 give the performance of MTR and LANCELOT. The objective function values of both methods are given in the second and third columns. The CPU times (in seconds) are given in the fourth and fifth columns. The iterations given in the sixth and seventh columns represent the total number of conjugate gradient iterations performed by LANCELOT and the total number of TR subproblems generated by MTR, respectively. The total numbers of function and gradient evaluations for both codes are given in the last two columns.

Based on the results of Tables 6 and 7, we now give some conclusions about the performance of the MTR and LANCELOT codes in terms of CPU time and the relative difference (rel. diff.) of the objective values obtained by MTR and LANCELOT. As applied to those test problems, MTR has:

i) better CPU time and better or close optimal value (i.e., rel. diff. $\leq$ 1.0e-5) for 60%

Table 4: The main test problems and some results of MTR

Problem	Var	Free	Bound	Fix	TrsIt	Trs	Ratio
ARGLINA	387	387	0	0	11	6	1.83
ARGLINB	387	387	0	0	25	21	1.19
ARGLINC	388	388	0	0	23	19	1.21
ARWHEAD	10000	10000	0	0	13	12	1.08
BDQRTIC	19999	19999	0	0	180	136	1.32
BRATU1D	13333	13331	0	2	6508	6021	1.08
BROWNAL	547	547	0	0	11	8	1.38
BRYBND	14288	14288	0	0	70	61	1.15
CHEBYQAD	316	0	316	0	383	342	1.12
COSINE	20000	20000	0	0	27	14	1.93
CURLY20	14295	14295	0	0	1597	1471	1.09
CVXBQP1	20000	0	20000	0	1200	1162	1.03
DIXMAANA	19998	19998	0	0	21	15	1.4
DIXMAANB	19998	19998	0	0	21	15	1.4
DIXMAANC	19998	19998	0	0	23	17	1.35
DIXMAAND	19998	19998	0	0	25	19	1.32
DIXMAANJ	19998	19998	0	0	626	617	1.01
DIXMAANK	19998	19998	0	0	419	413	1.01
DIXMAANL	19998	19998	0	0	450	444	1.01
EDENSCH	13333	13333	0	0	55	43	1.28
EIGENBLS	1122	1122	0	0	1695	1608	1.05
EIGENCLS	1122	1122	0	0	1559	1452	1.07
ENGVAL1	19999	19999	0	0	32	26	1.23
ERRINROS	50	50	0	0	819	772	1.06
FLETCBV2	13330	13330	0	0	0	0	0.00
FMINSRF2	19881	19881	0	0	846	841	1.01
FREUROTH	19999	19999	0	0	97	61	1.59
FMINSURF	19881	19881	0	0	2405	2362	1.02
GRIDGENA	6218	0	5560	658	111	98	1.13
HILBERTB	282	282	0	0	16	11	1.45
LIARWHD	19999	19999	0	0	43	32	1.34
LINVERSE	10001	5000	5001	0	110	85	1.29
LMINSURF	19881	19321	0	560	1672	1669	1.00
MANCINO	183	183	0	0	32	21	1.52
NLMSURF	19881	19321	0	560	3618	3579	1.01
NOBNDTOR	12544	6050	6050	444	413	346	1.19
NONDIA	16666	16666	0	0	19	18	1.06

Table 5: The main test problems and some results of MTR(cont'd)

Problem	Var	Free	Bound	Fix	TrsIt	Trs	Ratio
OBSTCLAE	12769	0	12321	448	296	254	1.17
OBSTCLBL	12769	0	12321	448	135	113	1.19
OBSTCLBM	12769	0	12321	448	139	120	1.16
OBSTCLBU	12769	0	12321	448	141	117	1.21
ODC	16900	16384	0	516	682	682	1.00
PENALTY1	19999	19999	0	0	151	116	1.30
PENALTY3	120	120	0	0	94	76	1.24
POWELLSG	20000	20000	0	0	39	31	1.26
POWER	20000	20000	0	0	779	740	1.05
PROBPENL	19999	0	19999	0	10	6	1.67
QRTQUAD	20000	10000	10000	0	104	36	2.89
SENSORS	199	199	0	0	51	42	1.21
SINQUAD	19999	19999	0	0	82	54	1.52
SPARSQUR	16666	16666	0	0	41	37	1.11
TOINTGSS	20000	20000	0	0	29	21	1.38
TORSION2	12544	0	12100	444	357	298	1.20
TORSION4	12544	0	12100	444	203	180	1.13
TORSION6	12544	0	12100	444	149	125	1.19
TORSIONB	12544	0	12100	444	290	269	1.08
TORSIOND	12544	0	12100	444	244	200	1.22
TORSIONF	12544	0	12100	444	143	123	1.16
TQUARTIC	19999	19999	0	0	29	21	1.38
VARDIM	19999	19999	0	0	101	96	1.05

of problems;

ii) better CPU time and worse optimal value (i.e., rel. diff. $>$ 1.0e-5) for 10% of problems;

iii) worse CPU time and better or close optimal value for 23% of problems;

iv) worse CPU time and worse optimal value for 7% of problems.

Based on the above comparison, we see that it is promising to solve large-scale problems with simple bound constraints using our approach.

Table 6: Comparison of the Two Methods on the main test problems

Problem	Obj Value		Time		Iter		Nfg	
Name	MTR	LAN	MTR	LAN	MTR	LAN	MTR	LAN
ARGLINA	3.870000e+02	3.870000e+02	0.78	38.48	6	3	14	14
ARGLINB	1.931252e+02	1.931252e+02	2.25	29.00	21	1	44	13
ARGLINC	1.951252e+02	1.951252e+02	2.05	28.89	19	2	40	13
ARWHEAD	0.000000e+00	0.000000e+00	5.52	7.86	12	1	26	12
BDQRTIC	8.008640e+04	8.008640e+04	82.48	21.89	136	17	274	40
BRATU1D	5.840393e+06	1.261025e+08	3600.00	3600.00	6021	4540	12044	16852
BROWNAL	2.572063e-11	2.615568e-11	0.98	81.35	8	5	18	14
BRYBND	1.325909e-12	9.511770e-14	26.00	22.94	60	54	122	78
CHEBYQAD	6.778464e-03	8.047224e-03	224.84	3600.00	342	2394	698	2103
COSINE	-2.000000e+04	-1.999900e+04	8.35	5.16	14	11	30	33
CURLY20	-1.434017e+06	-1.434021e+06	580.10	3600.00	1471	43104	2944	47
CVXBQP1	2.894382e-04	9.000450e+06	577.99	238.05	1162	9760	2338	14
DIXMAANA	1.000000e+00	1.000000e+00	8.13	7.83	15	21	32	34
DIXMAANB	1.000000e+00	1.000000e+00	8.46	9.14	15	17	32	40
DIXMAANC	1.000000e+00	1.000000e+00	9.51	9.79	17	20	36	42
DIXMAAND	1.000000e+00	1.000000e+00	10.57	14.96	19	33	40	66
DIXMAANJ	1.000001e+00	1.000000e+00	325.88	52.51	617	577	1236	88
DIXMAANK	1.000001e+00	1.000000e+00	217.59	46.26	413	389	828	87
DIXMAANL	1.000001e+00	1.000000e+00	233.88	38.11	444	354	890	84
EDENSCH	8.000128e+04	8.000128e+04	15.97	10.18	43	25	88	70
EIGENBLS	1.974164e-03	1.125255e-01	172.40	3600.00	1608	3424	3218	6914
EIGENCLS	2.454111e-02	7.176403e+02	158.73	3600.00	1452	3131	2906	9172
ENGVAL1	2.219933e+04	2.219933e+04	14.35	5.83	26	10	54	28
ERRINROS	3.990416e+01	3.990415e+01	0.83	0.17	772	110	1546	209
FLETCBV2	-5.001006e-01	-5.001006e-01	0.17	0.59	0	0	2	2
FMINSRF2	1.000002e+00	1.000000e+00	413.21	676.48	841	1454	1684	2341
FREUROTH	2.433123e+06	2.433123e+06	39.90	11.97	61	20	124	47
FMINSURF	1.000002e+00	7.466224e+00	1176.94	3600.00	2362	106	4726	211
GRIDGENA	2.352000e+04	2.352000e+04	16.89	10.70	98	126	210	48
HILBERTB	6.262020e-16	8.799383e-12	1.41	12.19	11	31	24	86
LIARWHD	2.130882e-13	5.932646e-14	17.03	13.34	32	43	66	64
LINVERSE	3.409000e+03	3.409000e+03	29.54	3600.00	85	12067	184	16626
LMINSURF	9.000355e+00	9.000000e+00	856.11	3059.60	1669	7650	3340	10168
MANCINO	2.117119e-16	9.787775e-20	17.58	22.76	21	12	44	45
NLMSURF	3.914874e+01	4.290681e+01	1760.28	3600.00	3579	7237	7160	11997
NOBNDTOR	-4.418497e-01	-4.418594e-01	130.18	32.54	346	402	706	72
NONDIA	1.344590e-15	3.812729e-19	7.93	2.63	18	4	38	12

Table 7: Comparison of the Two Methods on the main test problems(cont'd)

Problem	Obj Value		Time		Iter		Nfg	
Name	MTR	LAN	MTR	LAN	MTR	LAN	MTR	LAN
OBSTCLAE	1.894773e+00	1.894763e+00	105.41	389.08	254	6264	522	26
OBSTCLBL	7.285840e+00	7.285834e+00	45.48	165.98	113	2938	240	36
OBSTCLBM	7.285838e+00	7.285834e+00	48.09	291.40	120	4431	254	16
OBSTCLBU	7.285840e+00	7.285834e+00	47.26	58.55	117	1107	248	40
ODC	-1.137898e-02	-1.082613e-02	374.75	3600.00	682	6018	1366	14458
PENALTY1	1.985763e-01	1.985846e-01	61.19	1769.96	116	30	206	124
PENALTY3	9.998808e-04	7.704032e-02	5.63	3600.00	76	14981	154	25899
POWELLSG	2.023535e-07	3.498944e-05	13.52	4.90	31	19	64	40
POWER	2.723993e-07	1.546668e-08	305.59	1222.89	740	68	1482	80
PROBPENL	1.000000e-08	1.007191e-08	4.03	202.40	6	5	22	18
QRTQUAD	-5.375101e+10	-1.526995e+10	34.46	3600.00	36	7354	72	23
SENSORS	-8.336250e+03	-8.064563e+03	18.57	265.23	42	282	86	1008
SINQUAD	-1.040172e+08	1.509966e-05	34.91	208.92	54	781	110	586
SPARSQUR	9.098914e-10	3.009349e-06	17.47	19.87	37	68	76	46
TOINTGSS	1.000050e+01	1.000050e+01	11.79	8.34	21	23	44	48
TORSION2	-4.263058e-01	-4.263350e-01	117.85	340.90	298	5327	610	28
TORSION4	-1.212871e+00	-1.212881e+00	69.53	459.81	180	7990	374	20
TORSION6	-2.859436e+00	-2.859446e+00	49.44	474.08	125	9381	264	18
TORSIONB	-4.183918e-01	-4.184089e-01	106.21	267.73	269	3829	552	20
TORSIOND	-1.204444e+00	-1.204454e+00	82.87	497.59	200	8585	414	22
TORSIONF	-2.850759e+00	-2.850769e+00	50.39	485.76	123	9229	260	16
TQUARTIC	1.098962e-14	2.620727e-09	10.51	15.01	21	53	44	50
VARDIM	1.524712e-15	3.050031e-04	40.28	3600.00	96	2	194	108

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

One goal of this thesis has been to study weighted paths in semidefinite programming (SDP). We have studied the limiting behavior of weighted infeasible central paths for SDP associated with the SDP map $X^{1/2}SX^{1/2}$ under the assumption that the problem has a strictly complementary primal-dual optimal solution. We have also derived an error bound on the distance between a point lying in a certain neighborhood of the central path and the set of primal-dual optimal solutions. A natural but challenging extension for future research is to analyze the limiting behavior of SDP weighted central paths without the strict complementarity assumption. Another interesting area for research relevant to this topic is to explore whether there exists a standard primal-dual interior point SDP algorithm, which is of polynomial and superlinear convergence automatically (i.e., with no need to perform multiple centrality steps between two consecutive standard steps).

Another goal of this thesis has been to develop an efficient method for solving large-scale SDPs. We have provided a new approach for solving well-structured large-scale SDPs via a saddle point mirror-prox algorithm by exploiting sparsity structure and reformulating them into smooth convex-concave saddle point problems. Through a set of computational experimentations, we have been capable of solving very huge SDPs, which are far beyond of the scope of interior-point methods and can be also challenging for other first-order methods. Of course, the sparsity pattern of SDPs plays important role in our approach. An interesting question is how to optimize the sparsity pattern by performing some reformulations or transformations such that the performance of our approach could be improved. Another area for future research is to apply our method to solve some SDPs from real-world problems (e.g., structural design).

We have also developed a long-step primal-dual infeasible path-following algorithm for convex quadratic programming (CQP) whose search directions are computed by means of

a preconditioned iterative linear solver. The preconditioner that we used in the analysis is restricted to maximum weight basis preconditioner. However, it might be possible to extend our analysis to a family of preconditioners (e.g., partial-update and ellipsoid preconditioners). Another intriguing but challenging area for future research is to extend our results to SDPs or general convex programming.

Finally we have developed an efficient "nearly exact" type of method for solving the large-scale "low-rank" trust region subproblems. Through a large set of computational results, we have established good performance of our method for solving large-scale nonlinear programming problems with simple bound constraints. It would be interesting to extend our approach to solve the large-scale problems with general constraints.

REFERENCES

[1] Adler, I. and Monteiro, R., "Limiting behavior of the affine scaling continuous trajectories for linear programming problems," *Mathematical Programming*, vol. 50, pp. 29–51, 1991.

[2] Alizadeh, F., "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM Journal on Optimization*, vol. 5, no. 1, pp. 13–51, 1995.

[3] Anstreicher, K., "Linear programming in $\mathcal{O}(n^3/(\ln n)\mathrm{L})$ operations," *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 803–812, 1999.

[4] Asic, M. D., Kovacevic-Vujcic, V. V., and Radosavljevic-Nikolic, M. D., "A note on limiting behavior of the projective and the affine rescaling algorithms," in *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988* (Lagarias, J. C. and Todd, M. J., eds.), vol. 114 of *Contemporary Mathematics*, pp. 151–157, Providence, Rhode Island, USA: American Mathematical Society, 1990.

[5] Baryamureeba, V. and Steihaug, T., "On the convergence of an inexact primal-dual interior point method for linear programming," Tech. Rep. 188, Department of Informatics, University of Bergen, 2000.

[6] Baryamureeba, V., Steihaug, T., and Zhang, Y., "Properties of a class of preconditioners for weighted least squares problems," Tech. Rep. 16, Department of Computational and Applied Mathematics, Rice University, 1999.

[7] Bayer, D. A. and Lagarias, J. C., "The nonlinear geometry of linear programming, Part I: Affine and projective scaling trajectories," *Transactions of the American Mathematical Society*, vol. 314, no. 2, pp. 499–526, 1989.

[8] Ben-Tal, A. and Bendsœ, M., "A new method for optimal truss topology design," *SIAM Journal on Optimization*, vol. 3, pp. 322–358, 1993.

[9] Ben-Tal, A. and Nemirovski, A., *Lectures on Modern Convex Optimization: Analysis, Algorithms, Engineering Applications*. MPS-SIAM Series on Optimization, Philadelphia: SIAM, 2000.

[10] Ben-Tal, A. and Nemirovskii, A., "Potential reduction polynomial time method for truss topology design," *SIAM Journal on Optimization*, vol. 4, pp. 596–612, 1994.

[11] Bentler, P. and Woodward, J., "Inequalities among lower bounds to reliability: with applications to test construction and factor analysis," *Psychometrika*, vol. 45, pp. 249–267, 1980.

[12] Bergamaschi, L., Gondzio, J., and Zilli, G., "Preconditioning indefinite systems in interior point methods for optimization," *Computational Optimization and Applications*, vol. 28, no. 2, pp. 149–171, 2004.

[13] Bertsekas, A., *Nonlinear Programming*. Belmont, Massachusetts: Athena Scientific, second ed., 1999.

[14] Boman, E., Chen, D., Hendrickson, B., and Toledo, S., "Maximum-weight-basis preconditioners," *Numerical Linear Algebra and Applications*, vol. 11, pp. 695–721, 2004.

[15] Boyd, S., Ghaoui, E., Feron, E., and Balakrishnan, V., *Linear matrix inequalities in system and control theory*, vol. 15 of *Studies in Applied Mathematics*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1994.

[16] Breitfeld, M. G. and Shanno, D. F., "Preliminary computational experience with modified log-barrier functions for large-scale nonlinear programming," in *Large Scale Optimization: State of the Art* (Hager, W. W., Hearn, D. W., and Pardalos, P. M., eds.), pp. 45–67, Kluwer Academics Publishers, 1994.

[17] Breitfeld, M. G. and Shanno, D. F., "Computational experience with penalty-barrier methods for nonlinear programming," *Annals of Operations Research*, vol. 62, pp. 439–463, 1996.

[18] Bunch, J. R. and Parlett, B. N., "Direct methods for solving symmetric indefinite systems of linear equations," *SIAM Journal on Numerical Analysis*, vol. 8, pp. 639–655, 1971.

[19] Burer, S. and Monteiro, R., "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming, Series B*, vol. 95, pp. 329–357, 2003.

[20] Burer, S., Monteiro, R., and Zhang, Y., "Solving a class of semidefinite programs via nonlinear programming," *Mathematical Programming*, vol. 93, pp. 97–122, 2002.

[21] Carpenter, T. and Vanderbei, R., "Symmetric indefinite systems for interior-point methods," *Mathematical Programming*, vol. 58, pp. 1–32, 1993.

[22] Cline, A. K., Moler, C. B., Stewart, G. W., and Wilkinson, J. H., "An estimate for the condition number of a matrix," *SIAM Journal on Numerical Analysis*, vol. 16, pp. 368–375, 1979.

[23] Conn, A. R., Gould, N. I. M., and Toint, P. L., *LANCELOT: a Fortran package for large-scale nonlinear optimization*. Springer Series in Computational Mathematics, Heidelberg, New York: Springer Verlag, 1992.

[24] Conn, A. R., Gould, N. I. M., and Toint, P. L., *Trust-region methods*. Philadelphia, Pennsylvania, USA: SIAM Publications, 2000.

[25] Cullum, J., Donath, W., and Wolfe, P., "The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices," *Mathematical Programming Study*, vol. 3, 1975.

[26] da Cruz Neto, J., Ferreira, O., and Monteiro, R., "Asymptotic behavior of the central path for a special class of degenerate SDP problems," manuscript, School of ISyE, Georgia Tech, Atlanta, GA, 30332, USA, July 2003.

[27] de Klerk, E., Roos, C., and Terlaky, T., "Initialization in semidefinite programming via a self-dual,skew-symmetric embedding," *Operations Research Letters*, vol. 20, pp. 213–221, 1997.

[28] de Klerk, E., Roos, C., and Terlaky, T., "Infeasible-start semidefinite programming algorithms via self-dual embeddings," *Fields Institute Communications*, vol. 18, pp. 215–236, 1998.

[29] Dennis, J. E. and Mei, H. H. W., "Two new unconstrained optimization algorithms which use function and gradient values," *Journal of Optimization Theory and Application*, vol. 28, pp. 453–482, 1979.

[30] Donath, W. E. and Hoffman, A. J., "Algorithms for partitioning graphs and computer logic based on eigenvectors of connection matrices," *IBM Technical Disclosure Bulletin*, vol. 15, 1972.

[31] Donath, W. E. and Hoffman, A. J., "Lower bounds for the partitioning of graphs," *IBM J. Res. and Devel.*, vol. 17, 1973.

[32] Fletcher, R., *Practical Methods of Optimization, Unconstrained Optimization*, vol. 1. New York: John Wiley, 1980.

[33] Fletcher, R., "A nonlinear programming in statistics (educational testing)," *SIAM Journal on Scientific and Statistical Computing*, vol. 2, pp. 257–267, 1981.

[34] Fletcher, R., "Semidefinte matrix constraints in optimization," *SIAM Journal on Control and Optimization*, vol. 23, pp. 493–513, 1985.

[35] Fortin, C. and Wolkowicz, H., "A survey of the trust region subproblem within a semidefinite programming framework," Tech. Rep. CORR 2002-22, University of Waterloo, Waterloo, Canada, 2002.

[36] Freund, R. M., "Complexity of an algorithm for finding an approximate solution of a semidefinite program with no regularity condition," Working paper OR 302-94, Operations Research Center, Massachusetts Institute of Technology, Cambridge, December 1994.

[37] Freund, R., Jarre, F., and Mizuno, S., "Convergence of a class of inexact interior-point algorithms for linear programs," *Mathematics of Operations Research*, vol. 24, no. 1, pp. 50–71, 1999.

[38] Fukuda, M., Kojima, M., Murota, K., and Nakata, K., "Exploiting sparsity in semidefinite programming via matrix completion I: General framework," *SIAM Journal on Optimization*, vol. 11, pp. 647–674, 2001.

[39] G. A. Shultz, R. B. S. and Byrd, R. H., "A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties," *SIAM Journal on Numerical Analysis*, no. 22, pp. 47–67, 1985.

[40] Gay, D. M., "Computing optimal locally constrained steps," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 2, pp. 186–197, 1981.

[41] Goemans, M. and Williamson, D., ".878-approximation algorithms for MAX CUT and MAX 2SAT," *Proceedings of the Symposium of Theoretical Computer Science*, pp. 422–431, 1994.

[42] Goldfarb, D. and Scheinberg, K., "Interior point trajectories in semidefinite programming," *SIAM Journal on Optimization*, vol. 8, pp. 871–886, 1998.

[43] Gould, N. I. M., Orban, D., and Toint, P. L., "General CUTEr documentation," Tech. Rep. TR/PA/02/13, CERFACS, Toulouse, France, 2003.

[44] Graña Drummond and H. Y. Peterzil, L. M., "The central path in smooth convex semidefinite programs," *Optimization*, vol. 51, pp. 207–233, 2002.

[45] Greenbaum, A., *Iterative Methods for Solving Linear Systems*. SIAM, 1997.

[46] Grone, B., Johnson, C. R., de Sa, E. M., and Wolkowicz, H., "Positive definite completions of partial hermitian matrices," *Linear Algebra and its Applications*, vol. 58, pp. 109–124, 1984.

[47] Güler, O., "Limiting behavior of the weighted central paths in linear programming," *Mathematical Programming*, vol. 65, pp. 347–363, 1994.

[48] Halická, M., "Analytical properties of the central path at the boundary point in linear programming," *Mathematical Programming*, vol. 84, pp. 335–355, 1999.

[49] Halická, M., "Two simple proofs of analyticity of the central path in linear programming," *Operations Research Letters*, vol. 28, pp. 9–19, 2001.

[50] Halická, M., "Analyticity of the central path at the boundary point in semidefinite programming," *European Journal of Operational Research*, vol. 143, pp. 311–324, 2002.

[51] Halická, M., de Klerk, E., and Roos, C., "Limiting behavior of the central path in semidefinite optimization," preprint, Faculty of Technical Mathematics and Informatics, TU Delft, NL–2628 CD Delft, The Netherlands, June 2002.

[52] Halická, M., de Klerk, E., and Roos, C., "On the convergence of the central path in semidefinite optimization," *SIAM Journal on Optimization*, vol. 12, pp. 1090–1099, 2002.

[53] Hebden, M. D., "An algorithm for minimization using exact second derivatives," Tech. Rep. T. P. 515, Atomic Energy Research Establishment, Harwell, England, 1973.

[54] Helmberg, C. and Rendl, F., "A spectral bundle method for semidefinite programming," *SIAM Journal on Optimization*, vol. 10, pp. 673–696, 2000.

[55] Hertog, D., *Interior Point Approach to Linear, Quadratic and Convex Programming: Algorithms and Complexity*, vol. 277 of *Mathematics and its Applications*. Dordrecht: Kluwer Academic Publishers, 1994.

[56] Jarre, F., "An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices," *SIAM Journal on Control and Optimization*, vol. 31, pp. 1360–1377, 1993.

[57] Kamath, A. and Karmarkar, N., "A continuous approach to compute upper bounds in quadratic maximization problems with integer constraints," in *Recent Advances in Global Optimization* (Floudas, C. and Pardalos, P., eds.), pp. 125–140, Oxford, UK: Princeton University Press, 1992.

[58] Kojima, M., Megiddo, N., and Mizuno, S., "A primal-dual infeasible-interior-point algorithm for linear programming," *Mathematical Programming*, vol. 61, no. 3, pp. 263–280, 1993.

[59] Kojima, M., Megiddo, N., Noma, T., and Yoshise, A., *A unified approach to interior point algorithms for linear complementarity problems*, vol. 538 of *Lecture Notes in Computer Science*. Berlin, Germany: Springer Verlag, 1991.

[60] Kojima, M., Mizuno, S., and Noma, T., "Limiting behavior of trajectories by a continuation method for monotone complementarity problems," *Mathematics of Operations Research*, vol. 15, no. 4, pp. 662–675, 1990.

[61] Kojima, M., Mizuno, S., and Yoshise, A., "A primal-dual interior point algorithm for linear programming," in *Progress in Mathematical Programming: Interior-Point and Related Methods* (Megiddo, N., ed.), ch. 2, pp. 29–47, Springer-Verlag, 1989.

[62] Kojima, M., Shida, M., and Shindoh, S., "Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs," *Mathematical Programming*, vol. 80, pp. 129–160, 1998.

[63] Kojima, M., Shida, M., and Shindoh, S., "A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction," *SIAM Journal on Optimization*, vol. 9, pp. 444–465, 1999.

[64] Kojima, M., Shindoh, S., and Hara, S., "Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices," *SIAM Journal on Optimization*, vol. 7, pp. 86–125, 1997.

[65] Korzak, J., "Convergence analysis of inexact infeasible-interior-point algorithms for solving linear programming problems," *SIAM Journal on Optimization*, vol. 11, no. 1, pp. 133–148, 2000.

[66] Kovacevic-Vujcic, V. and Asic, M., "Stabilization of interior-point methods for linear programming," *Computational Optimization and Applications*, vol. 14, pp. 331–346, 1999.

[67] Lovász, L., "On the Shannon Capacity of a graph," *IEEE Transactions of Information Theory*, vol. IT-25(1), pp. 1–7, January 1979.

[68] Lovász, L. and Schrijver, A., "Cones of matrices and setfunctions, and 0-1 optimization," *SIAM Journal on Optimization*, vol. 1, pp. 166–190, 1991.

[69] Lu, Z. and Monteiro, R., "A note on the local convergence of predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction," To appear in SIAM Journal on Optimization.

[70] Lu, Z. and Monteiro, R., "Limiting behavior of the Alizadeh-Haeberly-Overton weighted paths in semidefinite programming," manuscript, School of ISyE, Georgia Tech, Atlanta, GA, 30332, USA, July 2003.

[71] Luenberger, D., *Linear and Nonlinear Programming.* Addison-Wesley, 1984.

[72] Luo, Z.-Q., Sturm, J. F., and Zhang, S., "Superlinear convergence of a symmetric primal-dual path-following algorithm for semidefinite programming," *SIAM Journal on Optimization*, vol. 8, pp. 59–81, 1998.

[73] McLinden, L., "An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem," *Pacific Journal of Mathematics*, vol. 88, pp. 101–161, 1980.

[74] McLinden, L., "The complementarity problem for maximal monotone multifunctions," in *Variational Inequalities and Complementarity Problems* (Cottle, R., Giannessi, F., and Lions, J.-L., eds.), pp. 251–270, New York: Wiley, 1980.

[75] Megiddo, N., "Pathways to the optimal set in linear programming," in *Progress in Mathematical Programming: Interior Point and Related Methods* (Megiddo, N., ed.), pp. 131–158, New York: Springer Verlag, 1989. Identical version in: *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986.

[76] Milnor, J., *Singular points of complex hypersurfaces.* Ann. Math. Stud., Princeton University Press, 1968.

[77] Mizuno, S. and Jarre, F., "Global and polynomial-time convergence of an infeasible-interior-point algorithm using inexact computation," *Mathematical Programming*, vol. 84, pp. 357–373, 1999.

[78] Monteiro, R., "Convergence and boundary behavior of the projective scaling trajectories for linear programming," *Mathematics of Operations Research*, vol. 16, no. 4, pp. 842–858, 1991.

[79] Monteiro, R., "First- and second-order methods for semidefinite programming," *Mathematical Programming*, vol. 97, pp. 209–244, 2003.

[80] Monteiro, R. and O'Neal, J., "Convergence analysis of a long-step primal-dual infeasible interior-point LP algorithm based on iterative linear solvers," tech. rep., Georgia Institute of Technology, 2003.

[81] Monteiro, R., O'Neal, J., and Tsuchiya, T., "Uniform boundedness of a preconditioned normal matrix used in interior point methods," *SIAM Journal on Optimization*, vol. 15, no. 1, pp. 96–100, 2004.

[82] Monteiro, R. and Pang, J.-S., "Properties of an interior-point mapping for mixed complementarity problems," *Mathematics of Operations Research*, vol. 21, pp. 629–654, 1996.

[83] Monteiro, R. and Pang, J.-S., "On two interior-point mappings for nonlinear semidefinite complementarity problems," *Mathematics of Operations Research*, vol. 23, pp. 39–60, 1998.

[84] Monteiro, R. and Tsuchiya, T., "Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem," *Mathematics of Operations Research*, vol. 21, pp. 793–814, 1996.

[85] Monteiro, R. and Tsuchiya, T., "Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming," *SIAM Journal on Optimization*, vol. 9, pp. 551–577, 1999.

[86] Monteiro, R. and Zanjácomo, P., "General interior-point maps and existence of weighted paths for nonlinear semidefinite complementarity problems," *Mathematics of Operations Research*, vol. 25, pp. 381–399, 2000.

[87] Monteiro, R. and Zhou, F., "On the existence and convergence of the central path for convex programming and some duality results," *Computational Optimization and Applications*, vol. 10, pp. 51–77, 1998.

[88] Moré, J. J., "Recent developments in algorithms and software for trust region methods," in *Mathematical Programming: The State of the Art*, pp. 258–287, Berlin, Germany: Springer-Verlag, 1983.

[89] Moré, J. J. and Sorensen, D. C., "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553–572, 1983.

[90] Nemirovski, A., "Prox-method with rate of convergence $o(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM Journal on Optimization*, vol. 15, pp. 229–251, 2004.

[91] Nesterov, Y., "Dual extrapolation and its applications for solving variational inequalities and related problems," Tech. Rep. CORE Discussion Paper 2003/68, September 2003.

[92] Nesterov, Y., "Smooth minimization of nonsmooth functions," Tech. Rep. CORE Discussion Paper 2003/12, February 2003. To appear in Mathematical Programming.

[93] Nesterov, Y. E., "Long-step strategies in interior-point primal-dual methods," *Mathematical Programming*, vol. 76, pp. 47–94, 1996.

[94] Nesterov, Y. E. and Nemirovskii, A. S., "Polynomial barrier methods in convex programming," *Ekonomika i Matem. Metody*, vol. 24, pp. 1084–1091, 1988. (In Russian; English transl. Matekon: Translations of Russian and East European Math. Economics.).

[95] Nesterov, Y. E. and Nemirovskii, A. S., "Self-concordant functions and polynomial time methods in convex programming," Book–Preprint, Central Economic &

Mathematical Institute, USSR Acad. Sci. Moscow, USSR, 1989. Published in Nesterov and Nemirovsky [97].

[96] Nesterov, Y. E. and Nemirovskii, A. S., "Optimization over positive semidefinite matrices: Mathematical background and user's manual," Technical report, Central Economic & Mathematical Institute, USSR Acad. Sci. Moscow, USSR, 1990.

[97] Nesterov, Y. E. and Nemirovskii, A. S., *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*. Philadelphia: Society for Industrial and Applied Mathematics, 1994.

[98] Nocedal, J. and Wright, S. J., *Numerical optimization*. New York, USA: Springer-Verlag, 1999.

[99] Oliveira, A. and Sorensen, D., "Computational experience with a preconditioner for interior point methods for linear programming," Tech. Rep. 28, Department of Computational and Applied Mathematics, Rice University, 1997.

[100] Overton, M. and Womersley, R., "Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices," *Mathematical Programming*, vol. 62, pp. 321–357, 1993.

[101] Polyak, R., "Modified barrier functions (theory and methods)," *Mathematical Programming*, vol. 54, pp. 177–222, 1992.

[102] Portugal, L., Resende, M., Veiga, G., and Judice, J., "A truncated primal-infeasible dual-feasible network interior point method," *Networks*, vol. 35, pp. 91–108, 2000.

[103] Potra, F. A. and Sheng, R., "Superlinear convergence of interior-point algorithms for semidefinite programming," Reports on Computational Mathematics 86, Department of Mathematics, The University of Iowa, Iowa City, Iowa, April 1996.

[104] Potra, F. A. and Sheng, R., "A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming," *SIAM Journal on Optimization*, vol. 8, pp. 1007–1028, 1998.

[105] Powell, M. J. D., "A hybrid method for nonlinear equations," in *Numerical Methods for Nonlinear Algebraic Equations* (Rabinowitz, P., ed.), New York: Gordon and Breach, 1970.

[106] Preiß and J. Stoer, M., "High-order long-step method for solving semidefinite linear complementarity problem," To appear in Control and Cybernetics.

[107] Preiß and J. Stoer, M., "Analysis of infeasible-interior-point paths arising with semidefinite linear complementarity problems," *Mathematical Programming*, vol. 99, pp. 499–520, 2004.

[108] Pukelsheim, F., *Optimal design of experiments*. New York: Wiley, 1993.

[109] Reinsch, C. H., "Smoothing by spline functions ii," *Numerical Mathematics*, no. 16, pp. 451–454, 1971.

[110] Renegar, J., "Condition numbers, the barrier method, and the conjugate-gradient method," *SIAM Journal on Optimization*, vol. 6, pp. 879–912, 1996.

[111] Resende, M. and Veiga, G., "An implementation of the dual affine scaling algorithm for minimum cost flow on bipartite uncapacitated networks," *SIAM Journal on Optimization*, vol. 3, pp. 516–537, 1993.

[112] Saigal, R., Vandenberghe, L., and Wolkowicz, H., *Handbook of Semidefinite Programming*. Boston-Dordrecht-London: Kluwer Academic Publishers, 2000.

[113] Shapiro, A., "Weighted minimum trace factor analysis," *Psychometrika*, vol. 47, pp. 243–264, 1982.

[114] Shapiro, A., "Minimum trace factor analysis," in *Encyclopedia of Statistical Sciences* (Kotz, S. and Johnson, N., eds.), vol. 5, New York: John Wiley and Sons, 1985.

[115] Shor, N., "Dual quadratic estimates in polynomial and Boolean programming," *Annals of Operations Research*, vol. 25, pp. 163–168, 1990.

[116] Sorensen, D. C., "Newton's method with a model trust region modification," *SIAM Journal on Numerical Analysis*, no. 19, pp. 404–426, 1982.

[117] Sporre, G. and Forsgren, A., "Characterization of the limit point of the central path in semidefinite programming," Technical Report TRITA-MAT-2002-OS12, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, June 2002.

[118] Steihaug, T., "The conjugate gradient method and trust regions in large-scale optimization," *SIAM Journal on Numerical Analysis*, no. 20, pp. 626–637, 1983.

[119] Stoer, J. and Wechs, M., "Infeasible-interior-point paths for sufficient linear complementarity problems," *Mathematical Programming*, vol. 83, pp. 403–423, 1998.

[120] Stoer, J. and Wechs, M., "On the analyticity properties of infeasible-interior point paths for monotone linear complementarity problems," *Numerical Mathematics*, vol. 81, pp. 631–645, 1999.

[121] Sturm, J. F., "Superlinear convergence of an algorithm for monotone linear complementarity problems, when no strictly complementary solution exists," *Mathematics of Operations Research*, vol. 24, pp. 72–94, 1999.

[122] Todd, M. J., "Semidefinite optimization," manuscript, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA, August 2001.

[123] Todd, M., Tunçel, L., and Ye, Y., "Probabilistic analysis of two complexity measures for linear programming problems," *Mathematical Programming A*, vol. 90, pp. 59–69, 2001.

[124] Vaidya, P., "Solving linear equations with symmetric diagonally dominant matrices by constructing good preconditioners," tech. rep., 1991. A talk based on the manuscript was presented at the IMA Workshop on Graph Theory and Sparse Matrix Computation, October 1991, Minneapolis.

[125] Vandenberghe, L. and Boyd, S., "Semidefinite programming," *SIAM Review*, vol. 38, pp. 49–95, 1996.

[126] Vavasis, S. and Ye, Y., "A primal-dual interior point method whose running time depends only on the constraint matrix," *Mathematical Programming A*, vol. 74, pp. 79–120, 1996.

[127] Watson, G., "Algorithms for minimum trace factor analysis," *SIAM Journal on Matrix Analysis and its Applications*, vol. 13, pp. 1039–1053, 1992.

[128] Wechs, M., "The analyticity of interior-point-paths at strictly complementary solutions of linear programs," *Optimization, Methods and Software*, vol. 9, pp. 209–243, 1998.

[129] Witzgall, C., Boggs, P. T., and Domich, P. D., "On the convergence behavior of trajectories for linear programming," in *Mathematical Developments Arising from Linear Programming: Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988* (Lagarias, J. C. and Todd, M. J., eds.), vol. 114 of *Contemporary Mathematics*, pp. 161–187, Providence, Rhode Island, USA: American Mathematical Society, 1990.

[130] Wright, S., *Primal-Dual Interior-Point Methods*. SIAM, 1997.

[131] Zhang, J. and Xu, C., "A class of indefinite dogleg path methods for unconstrained minimization," *SIAM Journal on Optimization*, vol. 9, no. 3, pp. 646–667, 1999.

[132] Zhang, Y., "On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem," *SIAM Journal on Optimization*, vol. 4, no. 1, pp. 208–227, 1994.

[133] Zhou, G. and Toh, K.-C., "Polynomiality of an inexact infeasible interior point algorithm for semidefinite programming," *Mathematical Programming*, vol. 99, pp. 261–282, 2004.

VITA

Zhaosong Lu was born on August 20, 1971 in Huaining County, Anhui, China. In September 1989, he enrolled at Anhui Normal University in Wuhu, China, and earned a B.S. degree in Mathematics Education in July 1993. After one year, he enrolled at Xi'an Jiaotong University in Xi'an, China, and earned a M.S. degree in Biomathematics in July 1997. In January 1999, he enrolled at University of Alabama in Tuscaloosa, Alabama, and earned a M.S. degree in Applied Mathematics in May 2000. The following August, he enrolled at Georgia Institute of Technology in Atlanta, Georgia, and completed a Ph.D. degree in Operations Research in early August 2005. In late August 2005, he joined the Department of Mathematics Sciences at Carnegie Mellon University in Pittsburgh, Pennsylvania as a Zeev Nehari visiting assistant professor.